This Page Is Inserted by IFW Operations and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents will not correct images, please do not report the images to the Image Problem Mailbox.

(19) World Intellectual Property Organization International Bureau



(43) International Publication Date 18 October 2001 (18.10.2001)

PCT

(10) International Publication Number WO 01/77350 A2

(51) International Patent Classification7:

C12N 15/79

(21) International Application Number: PCT/US01/11436

(22) International Filing Date: 4 April 2001 (04.04.2001)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data: 09/545,574

7 April 2000 (07.04.2000) US

(71) Applicant (for all designated States except US): LARGE SCALE BIOLOGY CORPORATION [US/US]; 3333 Vaca Valley Parkway, Vacaville, CA 95688 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PALMER, Kenneth, E. [ZA/US]; 707 West Monte Vista Avenue, Vacaville, CA 95688 (US). POGUE, Gregory, P. [US/US]; 419 Trillick Court, Vacaville, CA 95688 (US).

(74) Agent: HALLUIN, Albert, P.; Howrey Simon Arnold & White, 301 Ravenswood Avenue, Box 34, Menlo Park, CA 94025 (US). (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

 without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

A2

(54) Title: COMPOSITIONS AND METHODS FOR INHIBITING GENE EXPRESSION

(57) Abstract: The present invention provides an eukaryotic recombinant vector suited for bi-directional transcription of a transgene to yield both sense and antisense RNA transcripts of the transgene in an eukaryotic cell. The invention vectors are particularly suited for mediating gene silencing in a variety of biological systems. The present invention also provides host cells and transgenic plants comprising the invention vectors. Further provided by the invention are methods of inhibiting expression of an endogenous gene present in an eukaryotic cell. Also included is a method of identifying a biological function(s) of an endogenous gene of interest in an eukaryotic cell by selectively inhibiting the expression of the endogenous gene.

COMPOSITIONS AND METHODS FOR INHIBITING GENE EXPRESSION

5

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the priority benefit of U.S. Patent Application 09/545,574, filed April 7, 2000, pending, which is hereby incorporated herein by reference in its entirety.

STATEMENT OF RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH

15

Not applicable.

TECHNICAL FIELD

This invention is in the field of genetic analysis. Specifically, the invention relates to the generation of a eukaryotic vector that allows bi-directional transcription of a transgene to yield both sense and antisense RNA transcripts from the same transgene. The compositions and methods embodied in the present invention are particularly useful for targeted inhibition of gene expression in a eukaryotic cell.

25

30

20

BACKGROUND OF THE INVENTION

The structure and biological behavior of a cell is determined by the pattern of gene expression within that cell at a given time. Perturbations of gene expression have long been acknowledged to account for a vast number of diseases including, numerous forms of cancer, vascular diseases, neuronal and endocrine diseases.

Abnormal expression patterns, in form of amplification, deletion, gene rearrangements, and loss or gain of function mutations, are now known to lead to aberrant behavior of a disease cell. Aberrant gene expression has also been noted as a defense mechanism of certain organisms to ward off the threat of pathogens.

One of the major challenges of genetic engineering has been to regulate the expression of targeted genes that are implicated in a wide diversity of physiological responses. While overexpression of an exogenously introduced transgene in a eukaryotic cell is relatively straightforward, targeted inhibition of specific genes has been more difficult to achieve. Traditional approaches for suppressing gene expression, including site-directed gene disruption, antisense RNA or co-suppressor injection, require complex genetic manipulations or heavy dosages of suppressors that often exceeds the toxicity tolerance level of the host cell.

5

10

15

20

25

30

Recently, a new technique, "double-stranded RNA interference" has emerged in the study of gene silencing. Several research groups have demonstrated a marked inhibition of a specific nuclear gene expression in a wide range of eukaryotes by introduction into cells of dsRNA fragments that bear sequence homology with the nuclear gene. For instance, Fire et al. (1998) Nature 395: 854 reported the success of gene-specific interference in C. elegans that was mediated by ingested E. coli carrying a prokaryotic vector capable of producing both sense and antisense RNAs of the selected C. elgans genes. Misquitta et al. demonstrated the targeted disruption of nautilus gene in Drosophila melanogaster by injecting into the Drosophila embryo multiple copies of nautilus dsRNA. See Misquitta et al. (1999) PNAS U.S.A. 96:1451-1456. Studies by Ngô et al. (1998) Proc. Natl. Acad. of Sci. U.S.A., 96:1451-1456 confirmed that dsRNA interference also occurs in certain protozoan species. Earlier studies by Cogoni et al. and Hamilton et al. suggested that formation of dsRNA play a pivotal role in gene silencing in fungi Neurospora crassa and other plants. See Cogoni et al. (1999) Nature 399: 166-169; Hamilton et al. (1999) Science 286: 950-952; and Waterhouse et al. (1999) PNAS U.S.A. 95: 13959-13964. More recent investigations by Wargelius et al. revealed that this phenomenon is also conserved in vertebrates such as the zebrafish. Wargelius et al. Biochem. Biophys. Res. Commun. 263: 156-161.

Current techniques for achieving RNA mediated gene silencing include: (a) use of prokaryotic vectors capable of transcribing both sense and antisense RNA (Fire et al. (1998) *Nature* 395: 854; (b) *in vitro* transcription of individual strands of a selected gene followed by annealing the transcribed sense and antisense RNAs (see, e.g. Misquitta et al. (1999) *PNAS U.S.A.* 96:1451-1456); and possibly (c) viruses induced gene silencing (see, e.g. Angell et al. (1997) *EMBO Journal* 16:

PCT/US01/11436 WO 01/77350

3675-3684; Angell et al. (1999) Plant Journal 20: 357-362). However, these methods bear a number of intrinsic limitations. First, none of these methods employs gene delivery vehicles that are applicable for consistent and persistent inhibition of gene expression in a eukaryote. Second, these existing methods do not necessarily result in production of a substantially homogenous population of dsRNAs. Notably, the in vitro preparation of double-stranded RNAs by transcribing and annealing sense RNA transcripts to antisense transcripts is time consuming, labor intensive, and not amenable for mass production or high-throughput analyses.

10

5

Thus, there remains a considerable need for compositions and methods to effect dsRNA-mediated gene silencing. An ideal reagent would be a self-replicating vector that is (a) capable of autonomous replication and expression of a selected transgene in a eukaryotic cell; and (b) capable of yielding both sense and antisense RNA transcripts from the same transgene, so as to effect production of dsRNA transcripts in a eukaryotic host cell. The present invention satisfies these needs and provides related advantages as well.

15

SUMMARY OF THE INVENTION

20

25

30

A principal aspect of the present invention is the design of a eukaryotic recombinant vector to effect gene silencing in a eukaryotic cell that is susceptible to dsRNA-mediated reduction of gene expression. Such a vector allows bi-directional transcription of a transgene to yield both sense and antisense RNA transcripts of the same transgene in a eukaryotic cell. While not being bound to any one theory, the production of dsRNAs induces transcriptional and/or post-transcriptional gene silencing in the host cell. Accordingly, the present invention provides a recombinant vector having the following unique characteristics: it comprises a viral replicon having two overlapping transcription units arranged in an opposing orientation and flanking a transgene of interest, wherein the two overlapping transcription units yield both sense and antisense RNA transcripts from the same transgene fragment in a eukaryotic host cell.

In one aspect of this embodiment, each of the overlapping transcription units of the vector comprises a promoter and a terminator that are arranged in one of the configurations shown in Figure 2(a)-(d). The promoter can be constitutive or

inducible; it can be active in all tissues and cell types of an organism or operative only in selected tissues (i.e. tissue-specific).

In another aspect, the recombinant vector comprises a viral replicon that is derived from a DNA virus. Such DNA viruses can be selected from the group consisting of Geminivirus, Caulimoviridae, Badnaviridae, Circoviridae, Circinoviridae, Parvoviridae, Papovaviridae, Polyomaviridae, Adenoviridae, Herpesviridae, Poxviridae, Iridoviridae, Baculoviridae, Hepadnaviridae, Retroviridae, Gyrovirus, Nanovirus, and African Swine Fever virus.

In yet another aspect, the subject vector is capable of autonomous replication in a eukaryotic cell.

In still another aspect, the subject vector is capable of inhibiting expression of genes endogenous to a eukaryotic host cell. Non-limiting representative eukaryotic cells whose gene expression can be inhibited upon introduction of the subject vectors are fungi, yeast cells, plant cells, inset, avian, mammalian or other animal cells. Preferably, the vectors effect a reduced expression of an endogenous gene that is substantially homologous to the transgene contained in the overlapping transcription units of the vectors. More preferably, delivery of the vectors into a suitable host cell results in a phenotypic change of the host cell. In certain preferred embodiments, the endogenous gene is native to the host cell. The endogenous gene can also be heterologous to the host cell. In some embodiments, the endogenous gene is a pathogenic gene derived from one or more members of the group consisting of virus, bacterium, fungus, and protozoa. The transgene carried in the vector can be a nucleotide sequence that encodes a membrane protein, a cytosolic protein, a secreted protein, a nuclear protein, or a chaperon protein.

The present invention also provides host cells transformed with the invention vectors. The present invention further provides a transgenic plant comprising a eukaryotic recombinant vector of the present invention.

Also provided by the present invention is a kit for generating a doublestranded RNA transcript in a eukaryotic cell that contains the subject vectors in suitable packaging.

Further embodied in the present invention is a method of inhibiting expression of an endogenous gene present in a eukaryotic cell. The method involves: (a) providing a eukaryotic recombinant vector containing a transgene

4

10

5

15

20

25

30

that is substantially homologous to the endogenous gene; (b) introducing the eukaryotic recombinant vector into the eukaryotic cell; and (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene that is contained in the transcription units of the vector, and thereby inhibiting expression of the corresponding endogenous gene in the eukaryotic cell.

5

10

15

20

25

30

6.

Also included in the present invention is a method of identifying a biological function(s) of an endogenous gene of interest in a eukaryotic cell by selectively inhibiting the expression of the endogenous gene. The method comprises: (a) providing a eukaryotic recombinant vector containing a transgene that is substantially homologous to the endogenous gene; (b) introducing the eukaryotic recombinant vector of (a) into the eukaryotic cell; (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene contained in the eukaryotic recombinant vector and thereby inhibiting expression of the endogenous gene in the eukaryotic cell; and (d) determining one or more phenotypic changes in the eukaryotic cell that correlate with the inhibited expression of the endogenous gene, thereby identifying the biological function(s) of the endogenous gene in the eukaryotic cell. In essence, the subject methods allow the creation of a transient or more long-term gene-specific knock-out system for analyzing the biological function of any endogenous gene of interest.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic representation of the process for production of dsRNA transcripts by a subject vector containing two overlapping transcription units.

Figure 2 (a)-(d) depict four different configurations of the overlapping transcription units of the subject vectors.

Figure 3 is a schematic representation of an exemplary construct MSVLSB-

Figure 4 depicts the nucleotide sequence of the vector pMSVLSB-1 (SEQ ID NO:9) described in Examples 1-2.

Figure 5 depicts the nucleotide sequence of the vector pMSVLSB-2 (SEQ ID NO:10) described in Examples 1-2.

Figure 6 depicts the nucleotide sequence of the vector pMSVLSB-3 (SEQ ID NO:11) described in Examples 1-2.

Figure 7 depicts the nucleotide sequence of the vector pMSVLSB-4 (SEQ ID NO:12) described in Examples 1-2.

Figure 8 depicts the nucleotide sequence of the vector pMSVLSB-5 (SEQ ID NO:13) described in Examples 1-2.

Figure 9 depicts the nucleotide sequence of the vector pMSVLSB-6 (SEQ ID NO: 14) described in Examples 1-2.

MODES FOR CARRYING OUT THE INVENTION

Throughout this disclosure, various publications, patents and published patent specifications are referenced by an identifying citation. The disclosures of these publications, patents and published patent specifications are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

General Techniques:

5

10

15

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See, e.g., Matthews, PLANT VIROLOGY, 3rd edition (1991); Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A

LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and

As used in the specification and claims, the singular form "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a cell" includes a plurality of cells, including mixtures thereof.

Definitions:

5

10

15

20

25

30

A "plant cell" refers to the structural and physiological unit of plants, consisting of a protoplast and the cell wall.

A "protoplast" is an isolated cell without cell walls, having the potency for regeneration into cell culture, tissue or whole plant.

A "host cell" includes an individual cell or cell culture which can be or has been a recipient for vector(s) or for incorporation of nucleic acid molecules and/or proteins. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in genomic of total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation. A host cell includes cells transfected *in vivo* with a polynucleotide(s) of this invention.

The terms "polynucleotide", "nucleotides" and "oligonucleotides" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

A "gene" refers to a polynucleotide containing at least one open reading frame that is capable of encoding a particular protein after being transcribed and translated.

"Genes of a specific developmental origin" refer to genes expressed at certain but not all developmental stages. For instance, a gene may be of embryonic or adult origin depending on the stage during which the gene is expressed.

5

10

15

20

25

30

A "disease-associated" or "disease-causing" gene refers to any gene which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at normal or abnormal level.

A gene "database" denotes a set of stored data which represent a collection of sequences including nucleotide and peptide sequences, which in turn represent a collection of biological reference materials.

As used herein, "expression" refers to the process by which a polynucleotide is transcribed into mRNA and/or the process by which the transcribed mRNA (also referred to as "transcript") is subsequently being translated into peptides, polypeptides, or proteins. The transcripts and the encoded polypeptides are collectedly referred to as gene product. If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in an eukaryotic cell.

"Differentially expressed", as applied to nucleotide sequence or polypeptide sequence in a subject, refers to over-expression or under-expression of that sequence when compared to that detected in a control. Underexpression also encompasses absence of expression of a particular sequence as evidenced by the absence of detectable expression in a test subject when compared to a control.

"Differential expression" refers to alterations in the abundance or the expression pattern of a gene product.

A "primer" is a short polynucleotide, generally with a free 3' -OH group, that binds to a target or "template" potentially present in a sample of interest by hybridizing with the target, and thereafter promoting polymerization of a polynucleotide complementary to the target.

The term "hybridize" as applied to a polynucleotide refers to the ability of the polynucleotide to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues in a hybridization reaction. The hydrogen bonding may occur by Watson-Crick base pairing, Hoogstein binding, or in any other sequence-specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi-stranded complex, a single self-hybridizing strand, or any combination of these. The hybridization reaction may constitute a step in a more extensive process, such as the initiation of a PCR reaction, or the enzymatic cleavage of a polynucleotide by a ribozyme.

10

15

5

Hybridization can be performed under conditions of different "stringency". Relevant conditions include temperature, ionic strength, time of incubation, the presence of additional solutes in the reaction mixture such as formamide, and the washing procedure. Higher stringency conditions are those conditions, such as higher temperature and lower sodium ion concentration, which require higher minimum complementarity between hybridizing elements for a stable hybridization complex to form. In general, a low stringency hybridization reaction is carried out at about 40 °C in 10 x SSC or a solution of equivalent ionic strength/temperature. A moderate stringency hybridization is typically performed at about 50 °C in 6 x SSC, and a high stringency hybridization reaction is generally performed at about 60 °C in 1 x SSC.

20

25

When hybridization occurs in an antiparallel configuration between two single-stranded polynucleotides, the reaction is called "annealing" and those polynucleotides are described as "complementary". A double-stranded polynucleotide can be "complementary" or "homologous" to another polynucleotide, if hybridization can occur between one of the strands of the first polynucleotide and the second. "Complementarity" or "homology" (the degree that one polynucleotide is complementary with another) is quantifiable in terms of the proportion of bases in opposing strands that are expected to form hydrogen bonding with each other, according to generally accepted base-pairing rules.

30

In the context of polynucleotides, a "linear sequence" or a "sequence" is an order of nucleotides in a polynucleotide in a 5' to 3' direction in which residues that neighbor each other in the sequence are contiguous in the primary structure of the

polynucleotide. A "partial sequence" is a linear sequence of part of a polynucleotide which is known to comprise additional residues in one or both directions.

The terms "cytosolic", "nuclear" and "secreted" as applied to cellular proteins specify the extracellular and/or subcellular location in which the cellular protein is mostly localized. Certain proteins are "chaperons", capable of translocating back and forth between the cytosol and the nucleus of a cell.

5

10

.15

20

25

30

A "subject" as used herein refers to a biological entity containing expressed genetic materials. The biological entity is preferably can be plant, animal, or microorganisms including bacteria, viruses, fungi, and protozoa. Tissues, cells and their progeny of a biological entity obtained *in vivo* or cultured *in vitro* are also encompassed.

A "control" is an alternative subject or sample used in an experiment for comparison purpose. A control can be "positive" or "negative". For example, where the purpose of the experiment is to detect a differentially expressed transcript or polypeptide in cell or tissue affected by a disease of concern, it is generally preferable to use a positive control (a subject or a sample from a subject, exhibiting such differential expression and syndromes characteristic of that disease), and a negative control (a subject or a sample from a subject lacking the differential expression and clinical syndrome of that disease).

"Heterologous" means derived from a genotypically distinct entity from the rest of the entity to which it is being compared. For example, a promoter removed from its native coding sequence and operatively linked to a coding sequence other than the native sequence is a heterologous promoter.

A "cell line" or "cell culture" denotes bacterial, plant, insect or higher eukaryotic cells grown or maintained in vitro. The descendants of a cell may not be completely identical (either morphologically, genotypically, or phenotypically) to the parent cell.

A "vector" is a nucleic acid molecule, preferably self-replicating, which transfers an inserted nucleic acid molecule into and/or between host cells. The term includes vectors that function primarily for insertion of a DNA or RNA into a cell, replication of vectors that function primarily for the replication of DNA or RNA, and expression vectors that function for transcription and/or translation of the DNA

or RNA. Also included are vectors that provide more than one of the above functions.

An "expression vector" is a polynucleotide which, when introduced into an appropriate host cell, can be transcribed and translated into a polypeptide(s). An "expression system" usually connotes a suitable host cell comprised of an expression vector that can function to yield a desired expression product.

A "replicon" refers to a polynucleotide comprising an origin of replication (generally referred to as an <u>ori</u> sequence) which allows for replication of the polynucleotide in an appropriate host cell. Examples of replicons include episomes (such as plasmids), as well as chromosomes (such as the nuclear or mitochondrial chromosomes).

A "transcription unit" is a DNA segment capable of directing transcription of a gene or fragment thereof. Typically, a transcription unit comprises a promoter operably linked to a gene or a DNA fragment that is to be transcribed, and optionally regulatory sequences located either upstream or downstream of the initiation site or the termination site of the transcribed gene or fragment.

Vectors of the present invention

5

10

15

20

25

30

A central aspect of the present invention is the design of a recombinant vector suited for bi-directional transcription of a transgene to yield both sense and antisense RNA transcripts of the transgene in a eukaryotic cell. The invention vectors are particularly suited for mediating nuclear gene silencing in a variety of biological systems. Distinguished from the previously described DNA vectors, the subject vectors have the following unique characteristics: (a) the vector replicates and directs expression of a transgene in a eukaryotic cell; and (b) the vector comprises a replicon having two overlapping transcription units arranged in an opposing orientation and flanking a transgene of interest, wherein the two overlapping transcription units yield both sense and antisense RNA transcripts from the same transgene in a eukaryotic host cell.

Several factors apply to the design of vectors having the above-mentioned characteristics. First, the vector comprises a replicon having an origin of replication (generally referred to as an <u>ori</u> sequence) which permits replication of the vector in a eukaryotic host cell. A preferred replicon is one comprising viral sequences capable

of directing autonomous replication of the vector in an appropriate host cell. Non-limiting examples of viral replicons include sequences derived from DNA viruses such as Geminivirus, Caulimoviridae, Badnaviridae; Circoviridae, Circinoviridae, Parvoviridae, Papovaviridae, Polyomaviridae, Adenoviridae, Herpesviridae, Poxviridae, Iridoviridae, Baculoviridae, Hepadnaviridae, Retroviridae, Gyrovirus, Nanovirus, and African Swine Fever virus, or the like. In addition to the replication origin, a replicon typically carries a transcription unit that directs transcription of a transgene or a fragment thereof to yield a plurality of RNA transcripts.

5

10

15

20

25

30

A second consideration in designing the subject vector is to select two overlapping transcription units. By "overlapping" is meant that the two transcription units directs transcription of both DNA strands of the same transgene to yield a plurality of partially or perfectly double stranded RNA transcripts. The two overlapping transcription units are typically arranged in an opposing orientation so that each unit can drive transcription of one of the complementary strands from the same transgene, and thus facilitate the generation of double stranded RNA transcripts. Elements within a transcription unit include but are not limited to promoter regions, enhancer regions, repressor binding regions, transcription initiation sites, ribosome binding sites, translation initiation sites, protein encoding regions and introns, and termination sites for transcription and translation. Preferred transcription units are arranged in a configuration shown in Figure 2(a)-(d).

As used herein, a "promoter" is a DNA region capable under certain conditions of binding RNA polymerase and initiating transcription of a coding region located downstream (in the 3' direction) from the promoter. It can be constitutive or inducible. In general, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence is a transcription initiation site, as well as protein binding domains responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes.

The choice of promoters will largely depend on the host cells in which the vector is introduced. Commonly employed plant promoters include but are not limited those from agrobacterium, nopaline synthase gene, octopine synthase gene,

mannopine synthase, rbcS (small subunit of ribulose bis-phosphate carboxylase). In addition, the promoter sequences may be provided by viral material. Any RNA virus subgenomic promoters described in Dawson et al. Advances in Virus Research, 38:307-342 and WO93/03161 can thus be employed. For animal cells, a variety of robust promoters, both viral and non-viral promoters, are known in the art. Non-limiting representative viral promoters include CMV, the early and late promoters of SV40 virus, promoters of various types of adenoviruses (e.g. adenovirus 2) and adeno-associated viruses. It is also possible, and often desirable, to utilize promoters normally associated with a desired transgene sequence, provided that such control sequences are compatible with the host cell system. See Goeddel et al., Gene Expression Technology Methods in Enzymology Volume 185, Academic Press, San Diego, (1991), Ausubel et al, Protocols in Molecular Biology, Wiley Interscience (1994).

5

10

20

25

30

Suitable promoter sequences for other eukaryotic cells such as yeast cells include the promoters for 3-phosphoglycerate kinase, or other glycolytic enzymes, such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. Other promoters, which have the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization.

To optimize the yield of double-stranded RNAs formed from the sense and anti-sense strands transcribed by the overlapping units, it is preferable to use two promoters of comparable strength. The relative strength of the promoters can be determined or ascertained by any convention recombinant techniques and methods exemplified herein. Representative techniques are Northern blot hybridization and DNA array-based technologies. An illustrative promoter pair comprises MSV mp promoter and CaMV 35S RNA promoter.

Where desired, heterologous promoters that are removed from their native coding sequences and operatively linked to a transgene which it is not naturally

found linked, can be used in constructing the invention vectors. As such, any viral promoters described above can be used to drive the transcription of a non-viral transgenes; promoters of one class of genes can be employed to direct transcription of transgenes coding for other related or unrelated classes of proteins. In certain embodiments of the invention, it is preferable to employ inducible promoters to control the transcription of a transgene. A diverse variety of inducible promoters have been described in the art. Promoters of any endogenous genes whose expressions are inducible by internal or external factors can be employed. Factors applicable for transcription induction include but are not limited to hormones, heat shock, oxygen deficiency, light, stress and various chemicals. Commonly employed inducible promoters are β -gal promoter that is activated upon addition of IPTG; hps70 promoter that is inducible by heat shock; and ribulose-1,5-biphosphate carboxylase (RUBISCO) promoter that is regulated by light.

Tissue-specific promoters may also be used. A vast diversity of tissue specific promoters have been described and employed by artisans in the field. Representative plant tissue promoters include that of legumin (or other seed storage protein promoters), patatin and the like. Exemplary promoters operative in selective animal tissue include hepatocyte-specific promoters and cardiac muscle specific promoters. Depending on the intended use of the subject vectors, those skilled in the art will know of other suitable tissue-specific promoters applicable for non-constitutive bi-directional transcription.

In constructing the subject vectors, the termination sequences associated with the transgene are also inserted into the 3' end of the sequence desired to be transcribed to provide polyadenylation of the mRNA and/or transcriptional termination signal. The terminator sequence preferably contains one or more transcriptional termination sequences (such as polyadenylation sequences) and may also be lengthened by the inclusion of additional DNA sequence so as to further disrupt transcriptional read-through. Preferred terminator sequences (or termination sites) of the present invention have a gene that is followed by a transcription termination sequence, either its own termination sequence or a heterologous termination sequence. Examples of such termination sequences, including stop codons coupled to various polyadenylation sequences that are known in the art, widely available, and exemplified below. Where the terminator comprises a gene, it

can be advantageous to use a gene which encodes a detectable or selectable marker; thereby providing a means by which the presence and/or absence of the terminator sequence (and therefore the corresponding inactivation and/or activation of the transcription unit) can be detected and/or selected. Alternatively, a terminator may simply be a second promoter, arranged in inverted orientation to the promoter described above.

5

10

15

20

25

30

The terminators and promoters of the two overlapping transcription units may take a variety of configurations. In one aspect, terminators 1 and 2 of the overlapping transcription units are arranged to immediately flank the transgene as shown in Figure 2(a). In another aspect, the two terminators are placed at the 5' end or the 3' end of their respective promoters as depicted in Figure 2(b). In other aspects, terminator 1 and promoter 1 are flanked by terminator 2 and promoter 2 as shown in Figure 2(c), or vice versa (see Figure 2(d)). Any other variations in configuring the two overlapping transcription units that permit bi-directional transcription are encompassed by the present invention.

The transgene transcribed by an invention vector can be any gene expressed in a eukaryotic cell. The selection of transgene is determined largely by the intended purpose of the vector. Where the vector is used to inhibit expression of an endogenous gene present in a host cell, the transgene selected are substantially homologous to the target endogenous gene. In general, substantially homologous nucleotide sequences are at least about 60% identical with each other, after alignment of the homologous regions. Preferably, the sequences are at least about 75% identical; more preferably, they are at least about 80% identical; more preferably, they are at least about 90% identical; still more preferably, the sequences are 95% identical.

Sequence alignment and homology searches are often determined with the aid of computer methods. A variety of software programs are available in the art. Non-limiting examples of these programs are Blast (http://www.ncbi.nlm.nih.gov/BLAST/), Fasta (Genetics Computing Group package, Madison, Wisconsin), DNA Star, MegAlign, and GeneJocky. Any sequence databases that contains DNA sequences corresponding to a target gene or a segment thereof can be used for sequence analysis. Commonly employed databases include but are not limited to GenBank, EMBL, DDBJ, PDB, SWISS-PROT, EST,

STS, GSS, and HTGS. Sequence similarity can be discerned by aligning the transgene sequence against a target endogenous gene sequence. Common parameters for determining the extent of homology set forth by one or more of the aforementioned alignment programs include p value and percent sequence identity. P value is the probability that the alignment is produced by chance. For a single alignment, the p value can be calculated according to Karlin et al. (1990) Prco.Natl. Acad. Sci 87: 2264. For multiple alignments, the p value can be calculated using a heuristic approach such as the one programmed in Blast. Percent sequence identity is defined by the ratio of the number of nucleotide matches between the query sequence and the known sequence when the two are optimally aligned. A selected transgene and target endogenous sequences are considered to be substantially homologous when the regions of alignment exhibit the aforementioned range of percentage of identity using Fasta or Blast alignment program with the default settings.

15

5

10

Sequence homology can also be determined by functional analyses. A sequence that preserves the functionality of the nucleic acid with which it is being compared is particularly preferred. Functionality may be established by different criteria, such as ability to hybridize with a target polynucleotide, ability to effectively amplify a target sequence to yield a substantially homogenous multiplicity of products, and the ability to extend the 3' end sequence complementary to a target sequence in a nucleotide sequencing reaction.

20

25

Where desired, the transgene may comprise heterologous sequences that facilitate detection of the expression and purification of the gene product. Examples of such sequences are known in the art and include those encoding reporter proteins such as β-galactosidase, β-lactamase, chloramphenicol acetyltransferase (CAT), luciferase, green fluorescent protein (GFP) and their derivatives. Other heterologous sequences that facilitate purification may code for epitopes such as Myc, HA (derived from influenza virus hemagglutinin), His-6, FLAG, glutathione S-transferase (GST), maltose-binding protein (MBP), or the Fc portion of immunoglobulin.

30

The target endogenous genes whose expression is to be inhibited encompass native and heterologous genes present in the host cell. "Native" genes are nucleic acid sequences originated from the host cell. Non-limiting illustrative native genes

include those encode membrane proteins, cytosolic proteins, secreted proteins, nuclear proteins and chaperon proteins. Heterologous genes are sequences acquired exogenously by the host cell. Exogenous sequences can be either integrated into the host cell genome, or maintained as episomal sequences. An exemplary class of heterologous genes includes pathogenic genes derived from viruses, bacteria, fungi, and protozoa.

5

10

15

20

25

30

The endogenous genes suitable for the present invention may also be characterized based on one or more of the following features: ability to induce a phenotypic change in a host cell or organism, species origin, developmental origin, primary structural similarity, involvement in a particular biological process, association with or resistance to a particular disease or disease stage, tissue, subtissue or cell-specific expression pattern, and subcellular location of the expressed gene product. In one aspect, the endogenous gene may be any gene expressed in a eukaryote cell, such as a plant cell, animal cell or a yeast cell. In another aspect, the endogenous gene confers a phenotypic characteristic detectable by visual, microscopic, genetic, or chemical means. Within this class of genes, of particular interest are plant genes involved in growth phenotypes, e.g. stunting, hyperbranching, vein banding, ring spot, etching, and those responsible for color characteristics including bleaching and chlorosis. Also, of particular relevance are genes which upon inhibition provide an enhanced resistance to pathogens (e.g. bacteria, fungi, viruses, insects, and protozoa), and resistance to adverse environmental factors (e.g. temperature fluctuation, nutritional deficiency, adverse soil conditions, moisture, dryness, etc.).

In another aspect, the endogenous genes are of a specific developmental origin, such as those expressed in an embryo or an adult organism, during ectoderm, mesoderm, or endoderm formation in a multi-cellular animal, or during development of leaves, tubers, bud of a plant. In yet another aspect, the endogenous genes belong to a family of genes, or a sub-family of genes that share primary structural similarities. Structural similarities can be discerned with the aid of computer software described above. Non-limiting examples of gene families include those encoding proteinase, proteinase inhibitors, cell surface receptors, protein kinases (e.g. tyrosine, serine/threonine or histidine kinases), trimeric G-proteins, cytokines, PH-, SH2-, SH3-, PDZ-domain containing proteins, and any of those gene families

published by the Institute for Genomic Research (TIGR), Incyte Pharmaceuticals, Inc., Human Genome Sciences Inc., Monsanto, and PE-Celera.

5

10

-15

20

25

30

In yet another aspect, the endogenous genes are involved in a specific biological process, including but not limited to cell cycle regulation, cell differentiation, chemotaxsis, apoptosis, cell motility and cytoskeletal rearrangement. In still another aspect, the endogenous genes embodied in the invention are associated with a particular disease or with a specific disease stage. Such genes include but are not limited to those associated with autoimmune diseases, obesity, hypertension, diabetes, neuronal and/or muscular degenerative diseases, cardiac diseases, endocrine disorders, any combinations thereof. In yet still another aspect, the endogenous genes encompass those exhibiting restricted expression patterns. Non-limiting exemplary gene transcripts of this class include those that are not ubiquitously expressed, but rather are differentially expressed in one or more of the plant tissues including leaf, seed, tuber, stems, root, and bud; or expressed in animal body tissues including heart, liver, prostate, lung, kidney, bone marrow, blood, skin, bladder, brain, muscles, nerves, and selected tissues that are affected by various types of cancer (malignant or non-metastatic), affected by cystic fibrosis or polycystic kidney disease. Additional examples of non-ubiquitously expressed genes are those whose gene products are localized to certain subcellular locations: extracellular matrix, nucleus, cytoplasm, cytoskeleton, plasma and/or intracellular membranous structures which include but are not limited to coated pits, Golgi apparatus, endoplasmic reticulum, endosome, lysosome, and mitochondria.

In addition to the above-described elements, the vectors may contain a selectable marker (for example, a gene encoding a protein necessary for the survival or growth of a host cell transformed with the vector), although such a marker gene can be carried on another polynucleotide sequence co-introduced into the host cell. Only those host cells into which a selectable gene has been introduced will survive and/or grow under selective conditions. Typical selection genes encode protein(s) that (a) confer resistance to antibiotics or other toxins substances, e.g., ampicillin, neomycyin, methotrexate, etc.; (b) complement auxotrophic deficiencies; or (c) supply critical nutrients not available from complex media. The choice of the proper marker gene will depend on the host cell, and appropriate genes for different hosts are known in the art.

The vectors embodied in this invention can be obtained using recombinant cloning methods and/or by chemical synthesis. A vast number of recombinant cloning techniques such as PCR, restriction endonuclease digestion and ligation are well known in the art, and need not be described in detail herein. One of skill in the art can also use the sequence data provided herein or that in the public or proprietary databases to obtain a desired vector by any synthetic means available in the art.

Host cell and transgenic organisms of the present invention:

10

5

The invention provides eukaryotic host cells transformed with the recombinant DNA vectors described above. The recombinant vectors containing the transgene of interest can be introduced into a suitable eukaryotic cell by any of a number of appropriate means, including electroporation, transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; and infection (where the vector is coupled to an infectious agent). The choice of introducing vectors will often depend on features of the host cell.

15

20

25

30

For most animal cells, any of the above-mentioned methods is suitable for vector delivery. For plant cells, a variety of techniques derived from these general methods is available in the art. The host cells may be in the form of whole plants, isolated cells or protoplasts. Preferably, the cells are "intact" in that the cell comprises an outer layer of cell wall, typically composed of cellulose for protection and maintaining the rigidity of the plant cell. Illustrative procedures for introducing vectors into plant cells include Agrobacterium-mediated plant transformation, protoplast transformation, gene transfer into pollen, injection into reproductive organs and injection into immature embryos. As is evident to one skilled in the art, each of these methods has distinct advantages and disadvantages. Thus, one particular method of introducing genes into a particular plant species may not necessarily be the most effective for another plant species.

Agrobacterium tumefaciens-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, bypassing the need for regeneration of an intact plant from a protoplast. The use of Agrobacterium-mediated expression vectors to introduce

5

10

15

20

25

30

DNA into plant cells is well known in the art. This technique makes use of a common feature of Agrobacterium which colonizes plants by transferring a portion of their DNA (the T-DNA) into a host cell, where it becomes integrated into nuclear DNA. The T-DNA is defined by border sequences which are 25 base pairs long, and any DNA between these border sequences is transferred to the plant cells as well. The insertion of a recombinant plant viral nucleic acid between the T-DNA border sequences results in transfer of the recombinant plant viral nucleic acid to the plant cells, where the recombinant plant viral nucleic acid is replicated, and then spreads systemically through the plant. Agro-infection has been accomplished with potato spindle tuber viroid (PSTV); CaV; and Lazarowitz, S., Nucl. Acids Res. 16:229 (1988)) digitaria streak virus (Donson et al., Virology 162:248 (1988)), wheat dwarf and tomato golden mosaic virus (TGMV). Therefore, agro-infection of a susceptible plant could be accomplished with a virion containing a recombinant plant viral nucleic acid based on the nucleotide sequence of any of the above viruses. Particle bombardment or electrosporation or any other methods known in the art may also be used.

Because not all plants are natural hosts for Agrobacterium, alternative methods such as transformation of protoplasts may be employed to introduce the subject vectors into the host cells. For certain monocots, transformation of the plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments. See, for example, Potrykus et al., *Mol. Gen. Genet.*, 199:167-177 (1985); Fromm et al., *Nature*, 319:791 (1986); Callis et al., *Genes and Development*, 1:1183 (1987). Applicability of these techniques to different plant species may depend upon the feasibility to regenerate that particular plant species from protoplasts.

In addition to protoplast transformation, particle bombardment is an alternative and convenient technique for delivering the invention vectors into a plant host cell. Specifically, the plant cells may be bombarded with microparticles coated with a plurality of the subject vectors. Bombardment with DNA-coated microprojectiles has been successfully used to produce stable transformants in both plants and animals (see, for example, Sanford et al. (1993) Methods in Enzymology, 217:483-509). Microparticles suitable for introducing vectors into a plant cell are

typically made of metal, preferably tungsten or gold. These microparticles are available for example, from BioRad (e.g., Bio-Rad's PDS-1000/He). Those skilled in the art will know that the particle bombardment protocol can be optimized for any plant by varying parameters such as He pressure, quantity of coated particles, distance between the macrocarrier and the stopping screen and flying distance from the stopping screen to the target.

5

10

15.

20

30

Vectors can also be introduced into plants by direct DNA transfer into pollen as described by Zhou et al., Methods in Enzymology, 101:433 (1983); Luo et al., *Plant Mol. Biol. Reporter*, 6:165 (1988). Alternatively, the vectors can be injected into reproductive organs of a plant as described by Pena et al., *Nature*, 325:274 (1987).

Other techniques for introducing nucleic acids into a plant cell include:

- (a) Hand Inoculations. Hand inoculations are performed using a neutral pH, low molarity phosphate buffer, with the addition of celite or carborundum (usually about 1%). One to four drops of the preparation is put onto the upper surface of a leaf and gently rubbed.
- (b) Mechanized Inoculations of Plant Beds. Plant bed inoculations are performed by spraying (gas-propelled) the vector solution into a tractor-driven mower while cutting the leaves. Alternatively, the plant bed is mowed and the vector solution sprayed immediately onto the cut leaves.
- (c) High Pressure Spray of Single Leaves. Single plant inoculations can also be performed by spraying the leaves with a narrow, directed spray (50 psi, 6-12 inches from the leaf) containing approximately 1% carborundum in the buffered vector solution.
- 25 (d) Vacuum Infiltration. Inoculations may be accomplished by subjecting a host organism to a substantially vacuum pressure environment in order to facilitate infection.

Once introduced into a suitable host cell, expression of the transgene can be determined using any assay known in the art. For example, the presence of transcribed sense or anti-sense strands of the transgene can be detected and/or quantified by conventional hybridization assays (e.g. Northern blot analysis), amplification procedures (e.g. RT-PCR), SAGE (U.S. Patent No. 5,695,937), and

array-based technologies (see e.g. U.S. Pat. Nos. 5,405,783, 5,412,087 and 5,445,934). In conducting these analytical procedures, it is preferable to induce transcription of one strand of the transgene at a time. As is apparent to one skilled in the art, the simultaneous transcription of both sense and anti-sense strands facilitates formation of double stranded RNA molecules, which may obscure the accurate determination of the levels of sense and anti-sense RNA transcripts.

5

10

20

25

30

Expression of the transgene can also be determined by examining the protein product. A variety of techniques are available in the art for protein analysis. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunoflourescent assays, and PAGE-SDS.

In general, determining the protein level involves (a) providing a biological sample containing polypeptides; and (b) measuring the amount of any immunospecific binding that occurs between an antibody reactive to the transgene product and a component in the sample, in which the amount of immunospecific binding indicates the level of expressed proteins. Antibodies that specifically recognize and bind to the protein products of the transgene are required for immunoassays. These may be purchased from commercial vendors or generated and screened using methods well known in the art. See Harlow and Lane (1988) supra. and Sambrook et al. (1989) supra. The sample of test proteins can be prepared by homogenizing the eukaryotic transformants (e.g. plant cells) or their progenies made therefrom, and optionally solubilizing the test protein using detergents, preferably non-reducing detergents such as triton and digitonin. The binding reaction in which the test proteins are allowed to interact with the detecting antibodies may be performed in solution, or on a solid tissue sample, for example, using tissue sections or solid support that has been immobilized with the test proteins. The formation of the complex can be detected by a number of techniques known in the art. For example, the antibodies may be supplied with a label and unreacted antibodies may be removed from the complex; the amount of remaining label thereby indicating the amount of complex formed. Results obtained using any such assay on a sample

from a plant transformant or a progeny thereof is compared with those from a non-transformed source as a control.

5

10

20

25

30

The eukaryotic host cells of this invention are grown under favorable conditions to effect transcription of the transgene. Non-limiting examples of eukaryotic hosts are fungus, yeast, plant cells, insect, avian, mammalian or other animal cells. The host cells can be used, *inter alia*, as repositories of the transgene and/or vehicles for production of the transgene-specific double stranded RNAs. The host cells may also be employed to generate transgenic organisms such as transgenic animals and plants comprising the recombinant DNA vectors of the present invention. Preferred host cells are those having the propensity to regenerate into tissue or a whole organisms. Examples of these preferred host cells are oocytes, blastocytes, and certain plant cells exemplified herein.

Accordingly, this invention provides transgenic plants carrying the subject vectors. In a preferred embodiment, the trangenic plant exhibits a reduced expression (when compared to a control plant) of an endogenous gene that is substantially homologous to the transgene carried in the subject vector.

The regeneration of plants from either single plant protoplasts or various explants is well known in the art. See, for example, Methods for Plant Molecular Biology, Mary A. Shuler and Raymond E. Zielinski, Academic Press, Inc., San Diego, Calif. (1988). This regeneration and growth process includes the steps of selection of transformant cells and shoots, rooting the transformant shoots and growth of the plantlets in soil.

The regeneration of plants containing the subject vector introduced by Agrobacterium tumefaciens from leaf explants can be achieved as described by Horsch et al., Science, 227:1229-1231 (1985). In this procedure, transformants are grown in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant species being transformed as described by Fraley et al., Proc. Natl. Acad. Sci. U.S.A., 80:4803 (1983). This procedure typically produces shoots within two to four weeks and these transformant shoots are then transferred to an appropriate root-inducing medium containing the selective agent and an antibiotic to prevent bacterial growth. Transformant shoots that rooted in the presence of the selective agent to form plantlets are then transplanted to soil to allow

the production of roots. These procedures will vary depending upon the particular plant species employed, as is apparent to one of ordinary skill in the art.

A population of progeny can be produced from the first and second transformants of a plant species by methods well known in the art including cross fertilization and asexual reproduction. Transgenic plants embodied in the present invention are useful for production of desired proteins, and as test systems for analysis of the biological functions of a gene.

Uses of the vectors of the present invention:

10

5

The subject vectors provide specific reagents for inhibiting expression of an endogenous gene present in a host cell. The expression inhibition methods may be used in a wide variety of circumstances including suppression of a gene associated with a particular disease or disease stage; delineating the biological functions of a gene by analyzing a phenotypic change in the host cell that correlates with the selective suppression of gene expression; and facilitating drug screening by rendering the host cell more susceptible or resistant to a therapeutic agent of interest.

15

20

Accordingly, this invention provides a method of inhibiting expression of an endogenous gene present in a eukaryotic cell. The method comprises the steps of:

(a) providing a subject vector containing a transgene that is substantially homologous to an endogenous gene of a eukaryotic cell; (b) introducing the recombinant vector into the eukaryotic cell; (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene, and thereby inhibiting expression of the corresponding endogenous gene in the eukaryotic cell.

25

In a separate embodiment, the invention provides a method of identifying a biological function(s) of an endogenous gene of interest in a eukaryotic cell by selectively inhibiting the expression of the endogenous gene. The method involves:

(a) providing a recombinant vector of the present invention, wherein the transgene contained in the vector is substantially homologous to the endogenous gene; (b) introducing the recombinant vector of (a) into the eukaryotic cell; (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene contained in the recombinant vector and thereby inhibiting expression of the endogenous gene in the eukaryotic cell; and

30

(d) determining one or more phenotypic changes in the eukaryotic cell that correlate with the inhibited expression of the endogenous gene, thereby identifying the biological function(s) of the endogenous gene in the eukaryotic cell.

The host cells encompassed by these embodiments are eukaryotic cells susceptible to dsRNA-mediated "genetic interference". dsRNA induced gene silencing has been observed in a variety of multi-cellular organisms including but not limited to worms, fruitflies, protozoa, fungi, mammals, and zebrafish. Thus, cells from any of these exemplary organisms can be employed. Suitable host cells may be derived from primary cultures or subcultures generated by expansion and/or cloning of primary cultures. Any cells capable of growth in culture can be used as host cells. Of particular interest is the type of cell that differentially expresses (over-expresses or under-expresses) a disease-causing gene. As is apparent to one skilled in the art, various cell lines may be obtained from public or private repositories. The largest depository agent is American Type Culture Collection (http://www.atcc.org), which offers a diverse collection of well-characterized cell lines derived from a vast number of organisms and tissue samples.

10

15

20

25

30

Upon delivery of the subject vectors, the host cells are cultured under conditions favorable for gene transcription. The parameters governing eukaryotic cell survival are generally applicable for induction of gene transcription. The culture conditions are well established in the art. Physicochemical parameters which may be controlled in vitro are, e.g., pH, CO2, temperature, and osmolarity. The nutritional requirements of cells are usually provided in standard media formulations developed to provide an optimal environment. Nutrients can be divided into several categories: amino acids and their derivatives, carbohydrates, sugars, fatty acids, complex lipids, nucleic acid derivatives and vitamins. Apart from nutrients for maintaining cell metabolism, most cells also require one or more hormones from at least one of the following groups: steroids, prostaglandins, growth factors, pituitary hormones, and peptide hormones to survive or proliferate (Sato, G.H., et al. in "Growth of Cells in Hormonally Defined Media", Cold Spring Harbor Press, N.Y., 1982; Barnes and Sato (1980) Anal. Biochem., 102:255. Given the vast wealth of information on the nutrient requirements, medium conditions optimized for cell survival, one skilled in the art can readily fashion various culture conditions using

PCT/US01/11436 WO 01/77350

any one of the aforementioned methods and compositions, alone or in any combination.

The inhibition of expression of the endogenous gene sharing substantial sequence homology with the transgene carried in the vectors can be determined by assaying for a difference, between the host cell and the control cell, in the level of mRNA transcripts of the endogenous gene. Alternatively, a suppression in expression is determined by detecting a difference in the level of the polypeptide(s) encoded by the endogenous gene. A preferred method is to detect a phenotypic change resulting from the decrease in expression of the endogenous gene of interest.

10

5

In assaying for an alteration in mRNA level, nucleic acid contained in the host cells is first extracted according to standard methods in the art. For instance, mRNA can be isolated using various lytic enzymes or chemical solutions according to the procedures set forth in Sambrook et al. (1989), supra or extracted by nucleicacid-binding resins following the accompanying instructions provided by 15 manufacturers. The mRNA contained in the extracted nucleic acid sample is then detected by hybridization (e.g. Northern blot analysis) and/or amplification procedures according to methods widely known in the art or based on the methods exemplified herein.

20

Reduction in expression of the endogenous gene can also be determined by examining the protein product of the endogenous gene. A variety of techniques is available in the art for protein analysis. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunoflourescent assays, and SDS-PAGE. In addition, cell sorting analysis can be employed to detect cell surface antigens. Such analysis involves labeling target cells with antibodies coupled to a detectable agent, and then separating the labeled cells from the unlabeled ones in a cell sorter. A sophisticated cell separation method is fluorescence-activated cell sorting (FACS). Cells traveling in single file in a fine stream are passed through a laser beam, and the fluorescence of each cell bound by the fluorescently labeled antibodies is then

30

measured.

25

Antibodies that specifically recognize and bind to the protein products of interest are required for conducting the aforementioned protein analyses. These antibodies may be purchased from commercial vendors or generated and screened using methods well known in the art. See Harlow and Lane (1988) *supra*. and Sambrook et al. (1989) *supra*.

10

5

Inhibition of gene expression can also result in phenotypic change(s) in a host cell. As used herein, phenotypic change refers to any non-genotypic change that can be detected visually, or analyzed biochemically or genetically. The choice of detection methods will largely depend on the nature of the phenotypic characteristics that are under investigation. For instance, certain phenotypic features of a plant cell can be detected microscopically or macroscopically. These features include improved tolerance to herbicides, improved tolerance to extremes of heat or cold, drought, salinity or osmotic stress; improved resistance to pests (insects, nematodes or arachnids) or diseases (fungal, bacterial or viral), production of enzymes or secondary metabolites; male or female sterility; dwarfness; early maturity; improved yield, vigor, heterosis, nutritional qualities, flavor or processing properties, and the like. Other detectable phenotypic changes are morphological alterations including but not limited to stunting, hyperbranching, vein banding, ring spot, etching, and those responsible for color characteristics including bleaching and chlorosis.

20

15

For animal cells, detectable phenotypic changes may encompass alterations in cell cycle regulation, cell differentiation, apoptosis, chemotaxsis, cell motility and cytoskeletal rearrangement. Methods for detecting these phenotypic changes are well-established in the art and hence are not detailed herein.

25

30

Other phenotypic changes commonly observed in both plant and animal cells involve differential expression (over-expression or under-expression) of a particular protein due to the selective inhibition of the endogenous gene of interest.

Differential gene expression may be analyzed by any chemical means available in the art or those disclosed herein. As is also apparent to artisans, altering expression of one endogenous gene may lead to changes in gene expression profile of a host of genes mapped to the same or related signal transduction pathways. As used herein, "signal transduction" refers to the process by which stimulatory or inhibitory signals are transmitted into and within a cell to elicit an intracellular response. Any

5

10

15

20

25

30

fluctuation in intracellular response of a eukaryotic host cell is also considered as a type of phenotypic change.

Alteration in intracellular response is often determined with the aid of reporter molecules. For example, when examining a signaling cascade involving a fluctuation of intracellular pH condition, pH sensitive molecules such as fluorescent pH dyes can be used as the reporter molecules. In another example where the signaling pathway of a trimeric G_q protein is analyzed, calcium-sensitive fluorescent probes can be employed as reporters. As is apparent to artisans in the field of signal transduction, trimeric G_q protein is involved in a classic signaling pathway, in which activation of G_q stimulates hydrolysis of phosphoinositides by phospholipase C to generate two classes of well-characterized second messengers, namely, diacylglycerol and inositol phosphates. The latter stimulates the mobilization of calcium from intracellular stores, and thus resulting in a transient surge of intracellular calcium concentration, which is a readout measurable with a calcium-sensitive probe.

Another exemplary class of reporter molecules is a reporter gene operably linked to an inducible promoter that can be activated upon the stimulation or inhibition of a signaling pathway. Reporter proteins can also be linked with other proteins whose expression is dependent upon the stimulation or suppression of a given signaling cascade. Commonly employed reporter proteins can be easily detected by a colorimetric or fluorescent assay. Non-limiting examples of such reporter proteins include: β -galactosidase, β -lactamase, chloramphenicol acetyltransferase (CAT), luciferase, green fluorescent protein (GFP) and their derivatives. Those skilled in the art will know of other suitable reporter molecules for assaying changes in a specific signaling transduction readout, or will be able to ascertain such, using routine experimentation.

To discern inhibition of gene expression, one typically conducts a comparative analysis of the subject and appropriate controls. Preferably, a test includes a positive control sample exhibiting a decrease in gene expression and a negative control having an unaltered expression level. The selection of an appropriate control cell or tissue is dependent on the sample cell or tissue initially selected and its phenotype which is under investigation.

In one aspect, the invention methods can be employed to selectively inhibit expression of an endogenous gene that is native to the eukaryotic host cell. Such a gene may encode encodes a protein selected from the group consisting of a membrane protein, a cytosolic protein, a secreted protein, a nuclear protein and a chaperon protein. Of particular interests are endogenous genes that confer phenotypic changes as a result of inhibition of the expression and/or function of the endogenous genes. In another aspect within this embodiment, the endogenous gene is heterologous to the host cell. As used herein, heterologous genes are acquired exogenously by the host cell. Non-limiting examples of heterologous genes are those derived from virus, bacterium, fungus, and protozoa.

In a separate embodiment, the invention methods are used to identify a biological function(s) of an endogenous gene in a eukaryotic cell by examining a phenotypic change associated with the inhibition in its expression and thus loss of biological function. In essence, the subject methods allow the creation of a transient or more long-term gene-specific knock-out system for analyzing the biological function of any endogenous gene of interest.

Kits comprising the vectors of the present invention

5

10

15

20

25

30

The present invention also encompasses kits containing the vectors of this invention in suitable packaging. Kits embodied by this invention include those that allow generation of a double-stranded RNA transcript in a eukaryotic cell.

Each kit necessarily comprises the reagents which render the delivery of vectors into a eukaryotic host cell possible. The selection of reagents that facilitate delivery of the vectors may vary depending on the particular transfection or infection method used. The kits may also contain reagents useful for generating labeled polynucleotide probes or proteinaceous probes for detection of gene silencing. Each reagent can be supplied in a solid form or dissolved/suspended in a liquid buffer suitable for inventory storage, and later for exchange or addition into the reaction medium when the experiment is performed. Suitable packaging is provided. The kit can optionally provide additional components that are useful in the procedure. These optional components include, but are not limited to, buffers, capture reagents, developing reagents, labels, reacting surfaces, means for detection, control samples, instructions, and interpretive information. The kits can be

employed to generate eukaryotic cells whose endogenous genes are selectively inhibited, and transgenic organisms comprising these eukaryotic cells.

Further illustration of the development and use of vectors and assays according to this invention are provided in the Example section below. The examples are provided as a guide to a practitioner of ordinary skill in the art, and are not meant to be limiting in any way.

5

EXAMPLES

Example 1: Construction of recombinant vectors comprising two opposing transcription units

5

10

We have designed a recombinant vector construct useful for silencing nuclear genes in many of the agriculturally-important cereal crops. The vector comprises sequences derived from maize streak geminivirus, isolated MSV-Kom (genbank accession number AF003952, classification: Family Geminiviridae, genus Mastrevirus, species maize streak virus, designated MSV-Komatipoort. Maize streak virus has a broad host range that encompasses all agriculturally important cereal crops, including but not limited to corn, wheat, rice, barley, rye, sorghum and millet. The methods for construction of infectious geminiviruses are well known to those skilled in the art, and are described in European patent application 8687015.5 as well as in US Patent No. 5,569,597.

15

20

We have synthesized a 1618 base pair synthetic DNA that contains the MSV-Kom repA and repB, long intergenic region (LIR) and short intergenic region (SIR) and thus all sequences that are required for viral replication. Palmer et al.(1999) Archives of Virology 144:1345-1360. This fragment was cloned into the pZeRO-2 vector (Invitrogen) as an EcoRI-XbaI fragment, to create the plasmid pMSVLSB-1, the sequence of which is shown in Figure 4. A 171 base pair fragment containing the movement protein (mp) promoter of MSV-Kom is synthesised and cloned into the pZeRO-2 vector as an HindIII-EcoRI fragment to create pMSVLSB-2 (sequence shown in Figure 5). The ApaI fragment containing the mp promoter is inserted between the two ApaI sites in pMSVLSB-1, to create pMSVLSB-3 (sequence shown in Figure 6).

25

30

The cauliflower mosaic virus 35S RNA promoter (CaMV 35S promoter) sequence is amplified with a vector containing this sequence (pBI121, from Clontech) as template DNA, using the following PCR primers containing the following restriction sites (shown in italicized): *EcoRI* in CaMV35SF and *SalI* in CaMV35SR.

CaMV35SF:

TTTGAATTCGTCAACATGGTGGAGCAC (SEQ ID NO:1)

CaMV35SR:

TTTGTCGACGTCCTCTCCAAATGAAATGAAC (SEQ ID NO:2)

5

WO 01/77350

The CaMV 35S promoter PCR product yielded is digested with *Eco*RI and *Sal*I and the restricted fragments are purified.

The zeocin resistance gene is amplified by PCR with the vector pZeRO-1

(Invitrogen) as template, using the following primers containing the following restriction sites shown in italicized: SalI, PacI and NotI in ZeoF and XhoI, PacI and NotI in ZeoR:

ZeoF:

15

CCCGTCGACTTAATTAAGCGGCCGCGTTTACAATTTCGCCTGATGC (SEQ ID NO:3)

ZeoR:

CCCCTCGAGTTAATTAAGCGGCCGCCTCAAAAAGGATCTTCACCTA
G (SEQ ID NO:4)

The zeocin resistance gene product yielded is digested with XhoI and SalI and purified.

25

20

The nopaline synthase (nos) terminator sequence is amplified by PCR with the vector pBI121 (Clontech) as template, using the following primers, with restriction sites XhoI in nosF and SpeI in nosR italicized:

NosF:

30

TTTCTCGAGCGAATTTCCCCGATCGTTCAAAC (SEQ ID NO:5)

NosR:

TTTACTAGTCCCGATCTAGTAACATAGATGAC (SEQ ID NO:6)

The nos terminator product yielded is digested with XhoI and SpeI and purified.

5

The digested CaMV35S promoter, zeocin resistance gene and nos terminator sequences are ligated together with T4 DNA ligase. The ligated product is diluted 1:100 in sterile water and the whole ligation product is re-amplified with the CaMV35SF and nosR primers. The resulting PCR product is digested with *EcoRI* and *SpeI*, purified and ligated with pMSVLSB-3 that is pre-digested with *EcoRI* and *SpeI*. The ligation reaction is used to transform *E. coli* competent cells. Transformants are selected on Luria Agar plates containing both kanamycin (100 µg/ml) and zeocin (50 µg/ml) to select for colonies containing the CaMV35S promoter-zeocin resistance gene-nos terminator cassette inserted into pMSVLSB-3 (Figure 6 and SEQ ID NO:11). Colonies putatively containing the correct plasmid are chosen, plasmid DNA isolated and screened by digestion with *EcoRI* and *SpeI*. One plasmid designated pMSVLSB-4 (Figure 7 and SEQ ID NO:12) is selected.

15

20

25

10

One of the methods in the art of construction of infectious clones of geminivirus genomes is to clone tandemly duplicated sequences of the geminivirus genome, with at least the LIR duplicated. This allows the virus sequence to escape from the cloning vector in planta by a replicative release mechanism. The virus Rep protein is transiently expressed in transfected cells, and induces a nick at each of the stem loop sequences contained within the origin of replication in the LIR. Rolling circle replication is initiated at each nick point, and this results in release of a ssDNA copy of the virus replicon, which is circularized by the Rep protein, and which then replicates autonomously in the plant cell nucleus. The XbaI-SpeI fragment from pMSVLSB-3, containing the viral LIR and Rep genes is inserted into the unique SpeI site in pMSVLSB-4 to create pMSVLSB-5 (Figure 8 and SEQ ID NO:13). The zeocin resistance gene is deleted by digestion with NotI; the DNA is recircularized and used to transform E.coli to kanamycin resistance with a new vector, pMSVLSB-6 (Figure 9 and SEQ ID NO:14). When the vector is introduced into plant cells, a monomeric copy of the insert is released by replicative release (described above) and replicates autonomously as construct MSVLSB-6 in the nuclei of infected cells.

30

The restriction map of construct MSVLSB-6 is shown in Figure 3; this genetic construct possesses the following features: (a) the *rep* genes and origins of replication from maize streak geminivirus that are necessary and sufficient for the autonomous replication of the viral construct and its associated foreign DNA in the host plant cell; (b) two overlapping transcription units present in the DNA replicon. The two overlapping transcription units are arranged according to the configuration shown in Figure 2. With reference to Figure 2, "promoter 1" and "terminator 1" in MSVLSB-6 are the MSV mp promoter and transcription termination signals present in the SIR, respectively, and "promoter 2" and "terminator 2" are the CaMV 35S RNA promoter and nos terminator sequences, respectively. The two overlapping transcription units share three unique restriction sites (*Sal*I, *Pac*I and *Not*I) and one non-unique restriction site (*Xho*I) where foreign DNA may be inserted so that it may be transcribed by both promoters to yield at least a partially double stranded RNA duplex of the foreign DNA sequence.

15

10

5

Example 2: Use of recombinant vectors to inhibit or silence gene expression in cereal crops:

Application of pMSVLSB-6 in inhibition of Dwarfl gene expression in rice

20

25

30

The vector pMSVLSB-6 exemplified above can be employed to inhibit expression of any endogenous gene in a variety of plant host cells. By way of illustration, the rice gene *Dwarfl* is inhibited to duplicate known mutant phenotype using a pMSVLSB-6 containing a fragment of the coding sequence of *Dwarfl*. (Genbank accession number AB028602). The gene is amplified from cDNA isolated from rice seedlings. Primer sequences are designed to have homology with the published sequence of Dwarfl. Ashikari *et al.* (1999) *PNAS U.S.A.* 96:10284-10289. The primer sequences contain *Not*I restriction sites at their 5' ends. The PCR product is digested with *Not*I and cloned into the *Not*I site of pMSVLSB-6 to generate pMSVLSB-6::dwarfls and pMSVLSB-6::dwarfla, with the insert cloned in the sense and antisense orientation with respect to the MSV mp promoter, respectively. The *XbaI-SpeI* fragment from each of these plasmids is transferred into an *Agrobacterium* binary vector that is commonly used for rice transformation. This vector is used to transform electrocompetent *Agrobacterium* strain LBA4404

(Life Technologies). Agrobacterium cultures containing the appropriate plasmids are used in transformation of rice. Transgenic rice is generated by standard protocols (see, e.g. US Patent 5,591,616). The transgenic rice plants display similar phenotypes to the dwarf1 mutant described by Ashikari et al. (1999) supra: they are giberellin-insensitve, dwarfed in comparison with un-silenced transgenic controls, and having broad, dark green leaves, compact pannicles and short, round grains.

Application of pMSVLSB-6 in inhibition of phytoene desaturase expression in maize seedlings

10

15

25

30

5

The coding sequence for the maize phytoene desaturase gene (pds), having the Genbank accession number U37285, is amplified from cDNA made from RNA isolated from four-day-old maize seedlings, of the cultivar "Golden Cross Bantam". The primers used for amplification of this cDNA have the following sequences containing the PacI sites (italicized) at the 5' ends:

zeapds1330:

TTTTTAATTAAGGTCCGCCTGAATTCTCG (SEQ ID NO:7)

20

zeapds1873 TTTTTAATTAACGGCAAGGCTCACAGTTTG (SEQ ID NO:8)

PCR amplification with these primers and cDNA made from RNA isolated from maize seedlings yields a product of 565 base pairs, which is then digested with *PacI*. The progenitor plasmid to pMSVLSB-6, pMSVLSB-5 is digested with *XbaI* and *SpeI* to release the MSV and associated overlapping transcription unit sequences from the pZeRO-2 cloning vector as a single 4816 base pair fragment. This fragment is cloned into the *Agrobacterium* binary vector pBin19 (Genbank: U09365) digested with *XbaI* to yield pMSVLSB-7. The plasmid pMSVLSB-7 is digested with *PacI* and the *pds* PCR fragment is inserted into this position, generating plasmid pMSVLSB-7::pds1 (cloned in the sense orientation with respect to the MSV mp promoter) and pMSVLSB-7::pds2 (cloned in the antisense orientation with respect to the MSV mp promoter. These two plasmids are each

introduced into Agrobacterium strain C58C1(pMP90) (Koncz and Schell, 1985) by electroporation. The Agrobacterium containing the binary vector plasmids is grown overnight in Luria Bertani medium containing appropriate selective antibiotics. The bacterial suspension is loaded into a 100 µl Hamilton syringe and injected into three day old maize seedlings (cultivar Golden Cross Bantam) according to methods described by Escudero et al. (1994) in the chapter "Agroinfection" of The Maize Handbook, Freelings M, Walbot V (eds). Plants that are successfully agroinfected display a photobleaching phenotype on the first three leaves, similar to that induced by spraying the plants with the phytoene desaturase-inhibitor norfluorazon.

10

CLAIMS

What is claimed is:

1. A eukaryotic recombinant vector comprising a viral replicon having two overlapping transcription units arranged in an opposing orientation and flanking a transgene of interest, wherein the two overlapping transcription units yield both sense and antisense RNA transcripts from the same transgene in a eukaryotic host cell.

10

5

- 2. The eukaryotic recombinant vector of claim 1, wherein each of the overlapping transcription units comprises a promoter and a terminator.
- 3. The eukaryotic recombinant vector of claim 2, wherein the promoter is a constitutive promoter.

15

4. The eukaryotic recombinant vector of claim 2, wherein the promoter is an inducible promoter.

5. The eukaryotic recombinant vector of claim 2, wherein the promoter is a 20

tissue-specific promoter.

6. The eukaryotic recombinant vector of claim 1, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(a).

25

7. The eukaryotic recombinant vector of claim 1, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(b).

30

8. The eukaryotic recombinant vector of claim 1, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(c).

9. The eukaryotic recombinant vector of claim 1, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(d).

5 10. The eukaryotic recombinant vector of claim 1 that inhibits gene expression of the eukaryotic host cell.

10

15

20

25

- 11. The eukaryotic recombinant vector of claim 1, wherein the eukaryotic host cell is selected from the group consisting of fungus, yeast cell, plant cell and animal cell.
 - 12. The eukaryotic recombinant vector of claim 1 that inhibits expression of an endogenous gene present in the host cell, wherein the endogenous gene is substantially homologous to the transgene contained in the overlapping transcription units.
 - 13. The eukaryotic recombinant vector of claim 12, wherein the endogenous gene is native to the host cell.
 - 14. The eukaryotic recombinant vector of claim 12, wherein the endogenous gene is heterologous to the host cell.
 - 15. The eukaryotic recombinant vector of claim 12, wherein the endogenous gene is a pathogenic gene derived from one or more members of the group consisting of virus, bacterium, fungus, and protozoa.
 - 16. The eukaryotic recombinant vector of claim 1, wherein expression of the transgene to yield double-stranded RNA transcripts confers a phenotypic change in the eukaryotic host cell.
 - 17. The eukaryotic recombinant vector of claim 1, wherein the transgene encodes a protein selected from the group consisting of a membrane protein, a cytosolic protein, a secreted protein, a nuclear protein, and a chaperon protein.

18. The eukaryotic recombinant vector of claim 1 that is an autonomously replicating vector.

- 19. The eukaryotic recombinant vector of claim 1, wherein the viral replicon is derived from a DNA virus.
- 20. The eukaryotic recombinant vector of claim 19, wherein the DNA virus is selected from the group consisting of Geminivirus, Caulimoviridae, Badnaviridae, Circoviridae, Circinoviridae, Parvoviridae, Papovaviridae, Polyomaviridae, Adenoviridae, Herpesviridae, Poxviridae, Iridoviridae, Baculoviridae, Hepadnaviridae, Retroviridae, Gyrovirus, Nanovirus, and African Swine Fever virus.

15

10

5

- 21. A host cell transformed with a vector of claim 1 or 10.
- 22. The host cell of claim 21 that is a eukaryotic cell selected from the group consisting of fungus, yeast cell, plant cell and animal cell.

20

23. A transgenic plant comprising a eukaryotic recombinant vector of claim 1 or 10.

25

- 24. The transgenic plant of claim 23 exhibiting reduced expression of an endogenous gene that is substantially homologous to the transgene contained in the eukaryotic recombinant vector.
- 25. A kit for generating a double-stranded RNA transcript in a eukaryotic cell comprising a eukaryotic recombinant vector of claim 1 in suitable packaging.

- 26. A method of inhibiting expression of an endogenous gene present in a eukaryotic cell, comprising:
 - (a) providing a eukaryotic recombinant vector of claim 12;

(b) introducing the eukaryotic recombinant vector into the eukaryotic cell:

- (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene that is contained in the transcription units of the vector, and thereby inhibiting expression of the corresponding endogenous gene in the eukaryotic cell.
- 27. The method of claim 26, wherein the endogenous gene is native to the host cell.

5

15

20

25

- 28. The method of claim 26, wherein the endogenous gene is heterologous to the host cell.
- 29. The method of claim 26, wherein the endogenous gene is a pathogenic gene derived from one or more members of the group consisting of virus, bacterium, fungus, and protozoa.
 - 30. The method of claim 26, wherein inhibition of the endogenous gene confers a phenotypic change in the host cell.
 - 31. The method of claim 26, wherein the host eukaryotic cell is selected from the group consisting of fungus, yeast cell, plant cell, and animal cell.
 - 32. The method of claim 26, wherein the eukaryotic recombinant vector is an autonomously replicating vector.
 - 33. The method of claim 26, wherein the eukaryotic recombinant vector comprises a viral replicon derived from a DNA virus.
 - 34. The method of claim 26, wherein the DNA virus is selected from the group consisting of Geminivirus, Caulimoviridae, Badnaviridae; Circoviridae, Circinoviridae, Parvoviridae, Papovaviridae, Polyomaviridae, Adenoviridae,

WO 01/77350

5

10

15

20

25

30

Herpesviridae, Poxviridae, Iridoviridae, Baculoviridae, Hepadnaviridae, Retrovirida, Gyrovirus, Nanovirus, and African Swine Fever virus.

- 35. The method of claim 26, wherein the eukaryotic recombinant vector comprises two overlapping transcription units, wherein each transcription unit comprises a promoter and a terminator.
 - 36. The method of claim 26, wherein the promoter is a constitutive promoter.
 - 37. The method of claim 26, wherein the promoter is an inducible promoter.
 - 38. The method of claim 26, wherein the promoter is a tissue-specific promoter.
 - 39. The method of claim 35, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(a).
 - 40. The method of claim 35, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(b).
 - 41. The method of claim 35, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(c).
- 42. The method of claim 35, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(d).
 - 43. A method of identifying a biological function(s) of an endogenous gene of interest in a eukaryotic cell by selectively inhibiting the expression of the endogenous gene, the method comprising:
 - (a) providing a eukaryotic recombinant vector of claim 12;
 - (b) introducing the eukaryotic recombinant vector of (a) in to the eukaryotic cell;

(c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene contained in the eukaryotic recombinant vector and thereby inhibiting expression of the endogenous gene in the eukaryotic cell; and

5

(d) determining one or more phenotypic changes in the eukaryotic cell that correlate with the inhibited expression of the endogenous gene, thereby identifying the biological function(s) of the endogenous gene in the eukaryotic cell.

10

- 44. The method of claim 43, wherein the eukaryotic cell is selected from the group consisting of fungus, yeast cell, plant cell, and animal cell.
 - 45. The method of claim 43, wherein the eukaryotic cell is a plant cell.

- 15

46. The method of claim 43, wherein the eukaryotic cell is an animal cell.

Figure 1

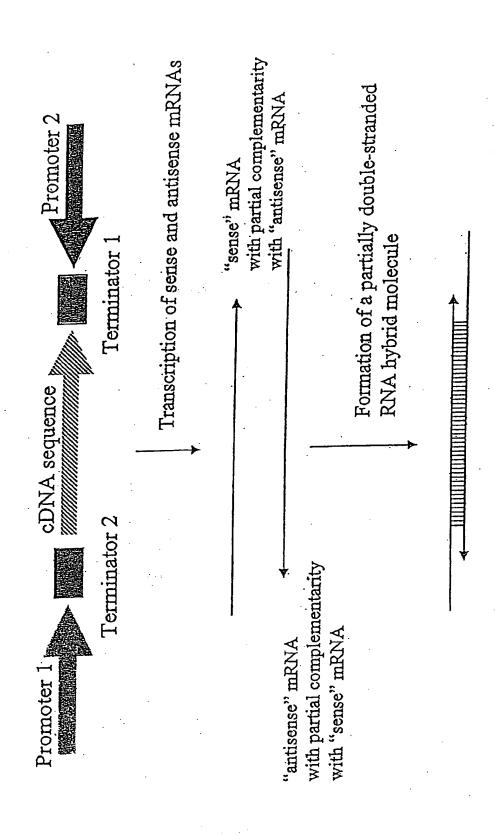


Figure 2

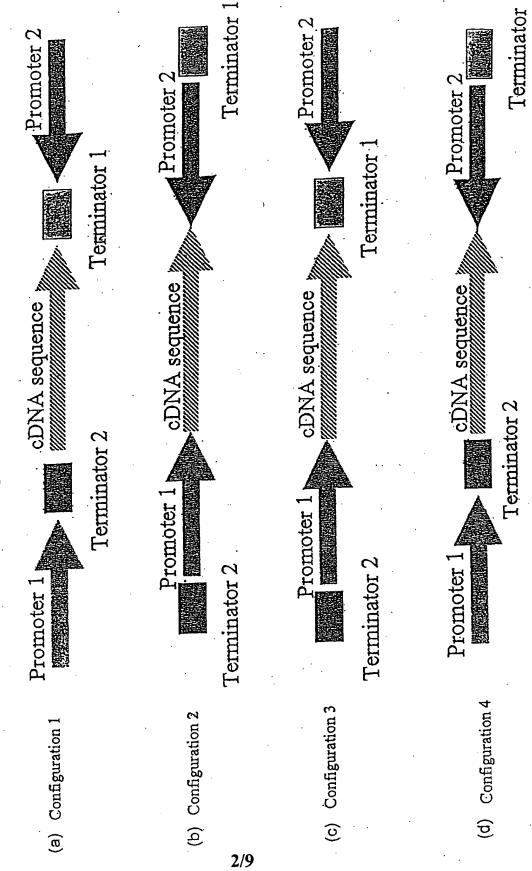
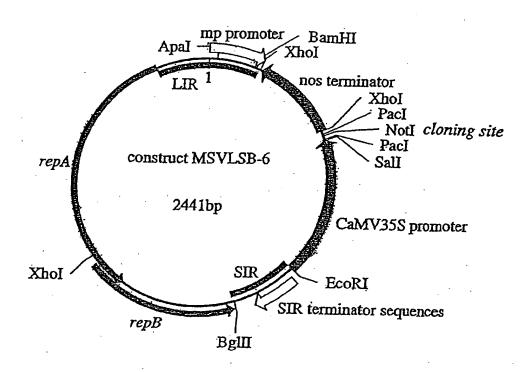


Figure 3



2101

2161

2221

2281

2341

2401

2461

2521

2581

2641

2701

2761

2821

2881

pMSVLSB-1: 4881 bp;

Figure 4

Composition 1161 A; 1260 C; 1251 G; 1209 T; 0 OTHER Percentage: 24% A; 26% C; 26% G; 25% T; 0%OTHER Molecular Weight (kDa) ssDNA: 1506.65 dsDNA: 3009.2 ORIGIN AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC 1 ACGACAGGIT ICCCGACIGG AAAGCGGGCA GIGAGCGCAA CGCAAITAAI GIGAGITAGC TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA 121 TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT 181 TTAGGTGACA CTATAGAATA CTCAAGCTAT GCATCAAGCT TGGTACCGAG CTCGGATCCA CTAGTAACGG CCGCCAGTGT GCTGGAATTC ATGGGCAGAC CCGTCTGTAC TTTAAGAGTG 301 TTGGCAACCA GTAATGAATA AAAACTCCCG TTTTATTATA TTTGATGAAT GCTGAAAGCT 361 TACATTAATA TGTCGTGCGA TGGCACGAAA AAACACACGC AAACAATACA GGGGGGTAGT 421 CGGCGGGCGG CTAAGGGTGG TGCTCGGCGG GCAGAACATC GAAAAATCAA GATCTATATG 481 AATTACACTT CCTCCGTAGG AGGAAGCACA GGGGGAGAAT ACCACTTCTC CCCCGGCGAC 541 ATAATGTAAA TGACGCAGTT TGCCTCGAAA TACTCCAGCT GCCCTGGAGT CATTTCCTTC 601 ATCCAATCTT CATCCGAGTT GGCGAGGATT ATTGTAGGCT TAGACTTCTT CTGCACCTTT 661 TICTICITAC CATACITGGG GTITACAATG AAATCCCTCT GACAGCCAAC TAACIGTITC 721 CAACAAGGAC AGAATTTAAA CGGAATATCA TCTACGATGT TGTAGATTGC GTCTTCGTTG 781 TATGAGACC AATCAACATT ATTTTGCCAG TAATTATGAA CCCCTAGGCT TCTGGCCCAA 841 GTAGATITIC CGGTTCTTGT TGGGCCGACG ATGTAGAGGC TCTGCTTTCT TGATCTTTCA 901 TCTGATGACT GGATACAGAA TCCATCCATT GGAGGTCAGA AATTGCATCC TCGAGGGTAT 961 ANCAGGTAGG TTGAAGGAGC ATGTAAGCTT CGGGACTAAC CTGGAAGATG TTAGGCTGGA 1021 GCCAATCGTT GATTGACTCA TTACAAAGTA AATCAGGTGA GGAGGGTGGA TGAGGATTGG 1081 TGAACTCTTC CTGAATCTCA GGAAAAAGCT TATTTGCAGA GTATTCAAAA TACTGCAATT 1141 TIGIGGACCA ATCANAGGGG AGCICITICI GGATCATGGA GAGGIACICI ICITIGGAGG 1201 TAGCGTGTGA AATAATGTCT CGCATTATTT CATCTTTAGA AGGCTTTTTT TCCTTTACCT 1261 CIGAATCAGA TITICCTAGG AAGGGGGACT TCCTAGGAAT GAAAGTACCT CTCTCAAACA 1321 CAGCCAGAGG TTCCTTGAGA ATGTAATCCC TCACTCTGTT AACTGACTTG GCACTCTGAA TATTTGGGTG AAACCCATTT ATATCAAAGA ACCTTGAGTC AGATATCCTT ATCGGCTTCT 1441 CTGGCTGAAG CAATGCATGT AAATGCAAAC TTCCATCTTT ATGTGCCTCT CGGGCACATA 1501 GANTATATTT GGGANTCCAN CGANCGACGA GCTCCCAGAT CATCTGACAG GCGATTTCAG 1561 GATTITCIGG ACACTITGGA TAGGITAGGA ACGTGITAGC GITCCIGIGI GAGAACIGAC 1621 GGTTGGATGA GGAGGAGGCC ATAGCCGACG ACGGAGGTTG AGGCTGAGGG ATGGCAGACT 1681 GGGAGCTCCA AACTCTATAG TATACCCGTG CGCCTTCGAA ATCCGCCGCT CCATTGTCTT 1741 ATAGTGGTTG TANATGGGCC GGACCGGGCC GGCCCAGCAG GAAAAGAAGG CGCGCACTAA 1801 TATTACCGCG CCTTCTTTC CTGCGAGGGC CCGGTAGGGA CCGAGCGCTT TGATTTAAAG 1861 CCTGGTTCTG CTTTGCGGCC GCTCGAGCAT GCATCTAGAG GGCCCAATTC GCCCTATAGT 1921 GAGTCGTATT ACAATTCACT GGCCGTCGTT TTACAACGTC GTGACTGGGA AAACCCTGGC 1981 GTTACCCAAC TTAATCGCCT TGCAGCACAT CCCCCTTTCG CCAGCTGGCG TAATAGCGAA 2041 GAGGCCCGCA CCGATCGCCC TTCCCCAACAG TTGCGCAGCC TATACGTACG GCAGTTTAAG

GTTTACACCT ATAAAAGAGA GAGCCGTTAT CGTCTGTTTG TGGATGTACA GAGTGATATT

ATTGACACGC CGGGGCGACG GATGGTGATC CCCCTGGCCA GTGCACGTCT GCTGTCAGAT

AAAGTCTCCC GTGAACTTTA CCCGGTGGTG CATATCGGGG ATGAAAGCTG GCGCATGATG

ACCACCGATA TGGCCAGTGT GCCGGTCTCC GTTATCGGGG AAGAAGTGGC TGATCTCAGC

CACCGCGAAA ATGACATCAA AAACGCCATT AACCTGATGT TCTGGGGAAT ATAAATGTCA

GECCTGAATG GCGAATGGAC GCGCCCTGTA GCGGCGCATT AAGCGCGCGG GTGTGGTGGT

TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT TCGCTTTCTT

CCCTTCCTTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC GGGGGCTCCC

TTTAGGGTTC CGATTTAGAG CTTTACGGCA CCTCGACCGC AAAAAACTTG ATTTGGGTGA

TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA CGTTGGAGTC

CACGITCITI AATAGIGGAC ICTIGITCCA AACIGGAACA ACACICAACC CIATCGCGGI

CTATTCTTT GATTTATAAG GGATGTTGCC GATTTCGGCC TATTGGTTAA AAAATGAGCT

GATTTAACAA AAATTTAAC AAAATTCAGA AGAACTCGTC AAGAAGGCGA TAGAAGGCGA

Figure 4 (cont'd)

				· ANACCACGAC	GAAGCGGTCA	GCCCATTCGC
2941	TGCGCTGCGA	ATCGGGAGCG	GCGATACCG	CCNACCCTAT	GTCCTGATAG	CGGTCCGCCA
3001	CGCCAAGCTC	TTCAGCAATA	TCACGGGIAG	, ANNACCECC	ATTTTCCACC	ATGATATICG
3061	CACCCAGCCG	GCCACAGTCG	ATGAATCCAG	CARCOCCOCC	GTCGGGCATG	CTCGCCTTGA
3121	GCAAGCAGGC	ATCGCCATGG	GTCACGACGA	GAICCICOCC	TTCGTCCAGA	TCATCCTGAT
3181	GCCTGGCGAA	CAGTICGGCI	GGCGCGAGCC	CCIGAIGCIC	GCGATGTTTC	GCTTGGTGGT
3241	CGACAAGACC	GGCTTCCATC	CGAGTACGIG	CICOCICAC	GCGATGTTTC	GCCATGATGG
3301	CGAATGGGCA	.GGTAGCCGGA	TCAAGCGTAI	DCDCCDGDCC	CTGCCCCGGC	ACTTCGCCCA
3361	ATACTITCTC	GGCAGGAGCA	AGGIGAGAIG	DASCINOCACA DASCINOCACA	CACAGCTGCG	CAAGGAACGC
3421	ATAGCAGCCA	GTCCCTTCCC	GCTTCAGTGA	CHACGICGAC	CAGTTCATTC	AGGGCACCGG
3481	CCCTCCTGCC	CAGCCACGAT	AGCLGCGCIG	CCICGICIA	TGACAGCCGG	AACACGGCGG
3541	ACAGGTCGGT	CTTGACAAAA	AGAACCGGGC	ACTUATION OF THE	GAATAGCCTC	TCCACCCAAG
3601	CATCAGAGCA	GCCGATTGTC	TGTTGTGCCC	TOTOM AND TOR	COGNANCIGAT	CCTCATCCTG
3,661	CGGCCGGAGA	ACCTGCGTGC	AATCCATCTT	GIICHAICAI	GCGAAACGAT	AAAGCCATCC
3721	TCTCTTGATC	AGATCTTGAT	CCCCIGCGCC	ATCAGATCCT	TGGCGGCGAG	AATTCCCGTT
3781	AGTITACTTT	GCAGGGCTTC	CCAACCITAC	CAGAGGGGGG	CCCAGCTGGC	CTGCAAGCTA
3841	CGCTTGCTGT	CCATAAAACC	·GCCCAGTCTA	GCTATCGCCA	TGTAAGCCCA	CTGACATTCA
3901	CCTGCTTTCT	CTTTGCGCTT	GCGTTTTCCC	TIGICCAGAI	AGCCCAGTAG	TAGGTTAAGA
3961	TCCGGGGTCA	GCACCGTTTC	TGCGGACTGG	CITICIACGI	GAAAAGGATC	Cyclication
4021	TCCTTTTTGA	TARTCTCATG	ACCAAAATCC	CTTAACGIGA	GTTTTCGTTC	CCCCTAATCT
4081		~~~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~	יוהו האנוענים אינוע אינו	. THIS AGAILU	TITITIO	COCOTANICA
4141		KKKKKKKKK	CCACCCTAC	CAGCGGTGGT	TIGITIGCCG	CULT CERTOTION
4201	TACCAACTCT	TTTTCCGAAG	GTAACTGGCT	TCAGCAGAGC	GCAGATACCA	CCTACATACC
4261	TTCTAGTGTA	GCCGTAGTTA	GGCCACCACT	TCAAGAACTC	TGTAGCACCG	CCIACAIACC
4321	TOGCTOTGCT	AATCCTGTTA	CCAGTGGCTG	CTGCCAGTGG	CGATAAGTCG	TOTALINGO
4381	GGTTGGACTC	AAGACGATAG	TTACCGGATA	AGGCGCAGCG	GTCGGGCTGA	CTACACCCTC
4441	CGTGCACACA	-GCCCAGCTTG	GAGCGAACGA	CCTACACCGA	ACTGAGATAC	CCCCTAACCG
4501			CONTRACTOR	CHAGAAAGGC	CONCROCTUT	CCCCIII
4561			AND AND COLOR	ACCOUNT CALLS		7007777
4621	ATAGTCCTGT	CGGGTTTCGE	CACCTCTGAC	TIGAGCGICG	ATTTTTGTGA	CHCCCCCTUTT
4681						
4741						
4801	TTACCGCCTT	TGAGTGAGCT	GATACCGCIC	GCCGCAGCCG	AACGACCGAG	COCHOCHMIL
4861	CAGTGAGCGA	GGAAGCGGAA	G			

Figure 5

pMSVLSB-2: 3413 bp; Composition 777 A; 950 C; 884 G; 802 T; 0 OTHER Percentage: 23% A; 28% C; 26% G; 23% T; 0%OTHER

dsDNA: 2104.2 Molécular Weight (kDa): ssDNA: 1052.40 ORIGIN AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC 1 ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC 61 TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA 121 TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT 181 TTAGGTGACA CTATAGAATA CTCAAGCTAT GCATCAAGCT TGGGCCCGGT AGGGACCGAG 241 CGCTTTGATT TARAGCCTGG TTCTGCTTTG TATGATTTAT CTARAGCAGC CCARTCTARA 301 GAAACCGGTC CCGGGCACTA TAAATTGCCT AACAAGTGCG ATTCATTCAT GGATCCTTTA 361 AACTCGAGTC TAGAGGGCCC GAATTCTGCA GATATCCATC ACACTGGCGG CCGCTCGAGC . 421 ATGCATCTAG, AGGGCCCAAT TCGCCCTATA GTGAGTCGTA TTACAATTCA CTGGCCGTCG 481 TITTACAACG TCGTGACTGG GAAAACCCTG GCGTTACCCA ACTTAATCGC CTTGCAGCAC 541 ATCCCCCTTT CGCCAGCTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC 601 AGTTGCGCAG CCTATACGTA CGGCAGTTTA AGGTTTACAC CTATAAAAGA GAGAGCCGTT 661 ATCGTCTGTT TGTGGATGTA CAGAGTGATA TTATTGACAC GCCGGGGCGA CGGATGGTGA 721 TCCCCCTGGC CAGTGCACGT CTGCTGTCAG ATAAAGTCTC CCGTGAACTT TACCCGGTGG 781 TGCATATCGG GGATGAAAGC TGGCGCATGA TGACCACCGA TATGGCCAGT GTGCCGGTCT 841 CCGTTATCGG GGAAGAAGTG GCTGATCTCA GCCACCGCGA AAATGACATC AAAAAACGCCA 901 TTAACCIGAT GTTCTGGGGA ATATAAATGT CAGGCCTGAA TGGCGAATGG ACGCGCCCTG 961 TAGCGGCGCA TTAAGCGCGC GGGTGTGGTG GTTACGCGCA GCGTGACCGC TACACTTGCC 1021 AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCCGGC 1081 TTTCCCCGTC AAGCTCTAAA TCGGGGGCTC CCTTTAGGGT TCCGATTTAG AGCTTTACGG 1141 CACCTCGACC GCAAAAAACT TGATTTGGGT GATGGTTCAC GTAGTGGGCC ATCGCCCTGA .1201 TAGACGGTTT TTCGCCCTTT GACGTTGGAG TCCACGTTCT TTAATAGTGG ACTCTTGTTC 1261 CARACTEGAR CARCACTCAR CCCTATCGCG GTCTATTCTT TTGATTTATA AGGGATGTTG 1321 CCGATTTCGG CCTATTGGTT AAAAAATGAG CTGATTTAAC AAAAATTTTA ACAAAATTCA GAAGAACTCG TCAAGAAGGC GATAGAAGGC GATGCGCTGC GAATCGGGAG CGGCGATACC 1441 GTAAAGCACG AGGAAGCGGT CAGCCCATTC GCCGCCAAGC TCTTCAGCAA TATCACGGGT 1501 AGCCAACGCT ATGTCCTGAT AGCGGTCCGC CACACCCAGC CGGCCACAGT CGATGAATCC 1561 AGAAAAGCGG CCATTTTCCA CCATGATATT CGGCAAGCAG GCATCGCCAT GGGTCACGAC 1621 GAGATCCTCG CCGTCGGGCA TGCTCGCCTT GAGCCTGGCG AACAGTTCGG CTGGCGCGAG 1681 CCCCTGATGC TCTTCGTCCA GATCATCCTG ATCGACAAGA CCGGCTTCCA TCCGAGTACG 1741 TECTCECTCE ATECGATETT TCGCTTGGTG GTCGAATGGG CAGGTAGCCG GATCAAGCGT 1801 ATGCAGCCGC CGCATTGCAT CAGCCATGAT GGATACTTTC TCGGCAGGAG CAAGGTGAGA 1861 TGACAGGAGA TCCTGCCCCG GCACTTCGCC CAATAGCAGC CAGTCCCTTC CCGCTTCAGT 1921 GACAACGTCG AGCACAGCTG CGCAAGGAAC GCCCGTCGTG GCCAGCCACG ATAGCGGCGC 1981 TGCCTCGTCT TGCAGTTCAT TCAGGGCACC GGACAGGTCG GTCTTGACAA AAAGAACCGG 2041 GCGCCCCTGC GCTGACAGCC GGAACACGGC GGCATCAGAG CAGCCGATTG TCTGTTGTGC 2101 CCAGTCATAG CCGAATAGCC TCTCCACCCA AGCGGCCGGA GAACCTGCGT GCAATCCATC 2161 TTGTTCAATC ATGCGAAACG ATCCTCATCC TGTCTCTTGA TCAGATCTTG ATCCCCTGCG 2221 CCATCAGATC CTTGGCGGCG AGAAAGCCAT CCAGTTTACT TTGCAGGGCT TCCCAACCTT 2281 ACCAGAGGGC GCCCCAGCTG GCAATTCCGG TTCGCTTGCT GTCCATAAAA CCGCCCAGTC 2341 TAGCTATCGC CATGTAAGCC CACTGCAAGC TACCTGCTTT CTCTTTGCGC TTGCGTTTTC 2401 CCTTGTCCAG ATAGCCCAGT AGCTGACATT CATCCGGGGT CAGCACCGTT TCTGCGGACT 2461 GGCTTTCTAC GTGAAAAGGA TCTAGGTGAA GATCCTTTTT GATAATCTCA TGACCAAAAT 2521 CCCTTAACGT GAGTTTTCGT TCCACTGAGC GTCAGACCCC GTAGAAAAGA TCAAAGGATC 2581 TTCTTGAGAT CCTTTTTTTC TGCGCGTAAT CTGCTGCTTG CAAACAAAAA AACCACCGCT 2641 ACCAGCGGTG GTTTGTTTGC CGGATCAAGA GCTACCAACT CTTTTTCCGA AGGTAACTGG 2701 CTTCAGCAGA GCGCAGATAC CAAATACTGT CCTTCTAGTG TAGCCGTAGT TAGGCCACCA 2761 CTTCAAGAAC TCTGTAGCAC CGCCTACATA CCTCGCTCTG CTAATCCTGT TACCAGTGGC 2821 TGCTGCCAGT GGCGATAAGT CGTGTCTTAC CGGGTTGGAC TCAAGACGAT AGTTACCGGA 2881

Figure 5 (control)

2941 TAAGGCGCAG CGGTCGGGCT GAACGGGGGG TTCGTGCACA CAGCCCAGCT TGGAGCG 3001 GACCTACACC GAACTGAGAT ACCTACAGCG TGAGCTATGA GAAAGCGCCA CGCTTCC 3061 AGGGAGAAAG GCGGACAGGT ATCCGGTAAG CGGCAGGGTC GGAACAGGA AGCCACCT 3121 GGAGCTTCCA GGGGGAAACG CCTGGTATCT TTATAGTCCT GTCGGGTTTC GCCACCT 3121 ACTTGAGCGT CGATTTTTGT GATGCTCGTC AGGGGGGCCG AGCCTATGGA AAAACGC

Figure 6

PMSVLSB-3:

pMSVLSB2 Apa fragment inserted: 4961 bp; Composition 1190 A; 1276 C; 1262 G; 1233 T; 0 OTHER Percentage: 24% A; 26% C; 25% G; 25% T; 0%OTHER

Molecular Weight (kDa): ssDNA: 1531.26 dsDNA: 3058.5 ORIGIN

MOTEC		, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,		•		
ORIGI		r acgcaaacco	. הייייייריייריייי	CGCGTTGGC	GATTCATTAA	TGCAGCTGGC
1 .			・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	GTGAGCGCA	I CGCAATTAAT	GIGAGITAGE
61	ACGACAGGT.	A GGCACCCCAG	COTTO	TTATGCTTC	GCTCGTATG	TTGTGTGGAA
121	TCACICATIO	ATAACAATTI	GCITIMONO	ACAGCTATGA	CCATGATTAC	GCCAAGCTAT
181	TIGIGAGEG	ATAACAATTI A CTATAGAATA	CACACAGGAA	CCATCAAGCT	TGGTACCGAG	CTCGGATCCA
241	TTAGGTGACA	CTATAGAATA CCGCCAGTGT	CICARGCIAI	ಶಿಷ್ಣಾದಿದ್ದಾರಿ	CCGTCTGTAC	TTTAAGAGTG
301	CTAGTAACG	GTAATGAATA	GCIGGAAIIC	STOCKOPOLICE STOCKOPOLICE	יים מבורם במירייר	GCTGAAAGCT
361	TIGGCAACC	GTAATGAATA	AAAACICCCG	TITIATIAT	משממשמממ ב	CCCCCCTAGT
421	TACATTAATA	GTAATGAATA A TGTCGTGCGA	TGGCACGAAA	AMACACACGC	ממשתמת ממחם	CDTCTATATG
481	CGGCGGGCGG	CTAAGGGTGG	TGCTCGGCGG	GCAGAACAIC	. AUGMENTERS	OCCCCCCCO C
541	AATTACACTT	CTARGGGIGG	AGGAAGCACA	ma COCCAGAAA	ACCACITCIC	CATTTCCTTC
601	ATAATGTAAA	TGACGCAGTT	TGCCTCGAAA	TACTCCAGCI	OCCUIONAGI.	CTTTTCCTTC
661	ATCCAATCTT	CATCCGAGIT	GGCGAGGATT	ATTGTAGGCT	THOMOTOTICIT	THE PROPERTY OF THE PROPERTY O
721	TTCTTCTTAC	CATCCGAGII	GTTTACAATG	AAATCCCTCT	ORTHOCKHOO!	CHAMACIOILIC
781	CAACAAGGAC	AGAATTTAAA	CGGAATATCA	TCTACGATGT	TGIMOMITOC	GICTICGTIO
841	TATGAAGACC	AATCAACATT	ATTTTGCCAG	TAATTATGAA	CCCCIMOGCI	TCIOGCCCCAT
901	GTAGATTITC	CGGTTCTTGT	TGGGCCGACG	ATGTAGAGGC	TOTALLICE	TOTALCETAT
961	TCTGATGACT	GGATACAGAA	TCCATCCATT	GGAGGTCAGA	MATIGUATUR	TCOMOGOTAL
1021	AACAGGTAGG	TTGAAGGAGC	ATGTAAGCTT	CGGGACTAAC	CIRCHARATA	TORGOCIOGA
1081	GCCAATCGTT	GATTGACTCA	TTACAAAGTA	AATCAGGIGA	GGAGGG16GA	TONGONITOG
1141			~~~ * * * * * * * * * * * * * * * * * *	TATTIMECAUA	GTVT1 CVVVV	TWOTOMBAT
1201	TIGIGGACCA	ATCAAAGGGG	AGCTCTTTCT	GGATCATGGA	POCOMMENT	ACCEPTANTO COLL
1261	TAGCGTGTGA	ATTAAAGGGG	CGCATTATTT	CATCTTTAGA	AGGCIIIII	TOCTTIACCE
1321			א א כיביביביב א רייוי	TEACHMAN	GWWGTWC	C1C1C1
1381				44 .Dt L 17. hb.	MALIGMLIAG.	OCUCTOTORS.
1441			**************************************	ACT TITLE ALTER	MINIMICCIA	MICOUCE
1501			יומתתייחות הת	T-11 T A 11 1 1 1	WIGIOCCICI	COCO CE 1
1561	THE RESERVE THE PERSON NAMED IN	~~~\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\	ריכיא א ריכיא ריכיא	GCTCCCAGAI	CHICIGHTHR	GCGWIII CO.
1621		עבולהותותות עיוע	TOCCTTACKIA	ACGIGITAGE	GTICCIGIGI	Chickness area
1681		~~~~~~~~~~	NUMBER CONTRACTOR	A(TAIAIALAITIG	MAGGETGWAGGA	Windows
1741	GGGAGCTCCA	AACTCTATAG	TATACCCGTG	CCCTTCGAA	ATCCGCCGCT	CCHIIGICII
1801		ma 3 3 m(つ(*)(*)(*)(*)(*)	CCNCCCCCCCC	GGCCCAGCAG	CAAAAAGAAGG	CGCGCACIAG
1861		CALMINIAL MINISTER OF	CICCONGCCC	CCCGTAGGGA	CCGAGCGCII	IGHIIANNG .
1921		Q TALL VALCE MALLEY	TOTATA	GCAGCCCAAT	CIMMONANC	CGG7 CCCGGG
1981	~~ ~~~~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~	መጣጣርምንእጣእእ	CTCCCTTTCA	TTCATGGATC	CITTAAACTC	GAGICIAGAG
2041	COCCA NUTC	へんへん なかり なんしん	CACTOTOTO	ACAATTCACT	GGCCGTCGTT	TIACAACGIC
2101	~~~~~~~~~	N N N CICCOTYCCC	ር ልግግን ለጥጥር	TTAATCGCCT	TGCAGCACAT	CCCCTITCG
2161	-c> comocoo	መእንመአመርመርእን	CACCCCCCA	CCGATCGCCC	TICCCAACAG	TIGUGUAGUU
2221	mama coma co	OUNCESTANCE.	ער איייטייט איייטייט איייטייט	ATAAAAGAGA	GAGCCGTTAT	CGICIGITIG
2281		CA COCA S DISTRAT	カツツマカ へんへんじ	CGGGGCGACG	GATGGTGATC	CCCTGGCCA
2341	CONTRACTOR CONTRACTOR	COCCOCACAT	あるなななですででで	GTGAACTTTA	CCCGGTGGTG	CATATCGGGG
2401	> ma> > > a a a a a	COCCAMCATC	カククカククスカイス '	TGGCCAGTGT	GCCGGTCTCC	GIIMICGGG
2461		DOMESTIC PROPERTY.	しょしいこうしょう アフ	ATGACATCAA	AAACGCCATT.	WWCCIONIOI
	mamadada a M	אליצויי איז א א א א א א א	ርርርርጥር እስጥና . !	GCGAATGGAU	GCGCCCTGTA	GCGGCGCGC
2521			かれたたただけるほと	RIGACCUCTA	CACTIBLEMO	COCCALIONS
2581	and a composition	MANAGEMENT OF THE PROPERTY OF	الململك كلملمك	CTCGCCACGT.	TUGUUGUTT	TCCCCGTCGT
2641		AAAAAAAAAAA I	المناسكات كالمستعلا	TTATTTAGAG	CITIMOGGOW	CCICONCOO
2701	PYYYYYOUNG	ATTTGGGTGA	TTROCOTTO	AGTGGGCCAT	CGCCCTGATA	GACGGTTTTT
2761	WHANNICI 10	WITT GOOT GW				

Figure 6 (cont'd)

			ON CONTRACTOR	AATAGTGGAC	TCTTGTTCCA	AACTGGAACA
2821	CGCCCTTTGA	CGTTGGAGTC	CHCGIICIII	AATAGTGGAC GATTTATAAG	GGATGTTGCC	GATITCGGCC
2881	ACACTCAACC	CTATCGCGGT	CIMITCITI	AAATTTAAC ATCGGAGCG	AAAATTCAGA	AGAACTCGTC
2941	TATTGGTTAA	AAAATGAGCT	MCCCCCCCC	ATCGGGAGCG TTCAGCAATA	GCGATACCGT	AAAGCACGAG
3001	AAGAAGGCGA	TAGAAGGCGA	TGCGC1GCW:	TTCAGCAATA	TCACGGGTAG	CCAACGCTAT
3061	GAAGCGGTCA	GCCCATTCGC	CACCAGCCG	GCCACAGTCG ATCGCCATGG	ATGAATCCAG	AAAAGCGGCC
3121	GTCCTGATAG	CGGTCCGCCA	CCAAGCAGGC	ATCGCCATGG CAGTTCGGCT	GTCACGACGA	GATCCTCGCC
318İ	ATTITCCACC	ATGATATICG	CCCTGGCGAA	CAGTTCGGCT	GGCGCGAGCC	CCTGATGCTC
3241	GTCGGGCATG	CTCGCCTIGA	CCACAAGACC	GGCTTCCATC	CGAGTACGTG	CTCGCTCGAT
3301	TTCGTCCAGA	TCATCCIGAT	CEANTGGGCA	GGTAGCCGGA GCCAGGAGCA	TCAAGCGTAT	GCAGCCGCCG
3361	GCGATGTTTC	GCITGGIGGI	BTACTTTCTC	GGCAGGAGCA GCCCTTCCC	AGGTGAGATG	ACAGGAGATC
3421	CATTGCATCA	GCCATGATGG	ATAGCAGCCA	GTCCCTTCCC CAGCCACGAT	GCTTCAGTGA	CAACGTCGAG
3481	CIGCCCCGGC	ACTICGCCCA	CCGTCGTGGC	CAGCCACGAT	AGCCGCGCTG	CCTCGTCTTG
3541	CACAGCIGCG	CAAGGAACGC	ACAGGTCGGT	CTTGACAAAA	AGAACCGGGC	GCCCCTGCGC
3601	CAGTTCATTC	AGGGCACCGG	CATCAGAGCA	GCCGATTGTC ACCTGCGTGC	TGTTGTGCCC	AGTCATAGCC
3661	TGACAGCCGG	ARCACGGCGG	CCCCGGAGA	ACCTGCGTGC AGATCTTGAT	AATCCATCTI	GTTCAATCAT
3721	GAATAGCCTC	TURCUCARG	TOTOTTGATC	AGATCTTGAT CCAGGGCTTC	CCCCTGCGCC	ATCAGATCCT
3781	GCGAAACGAT	222CCZATCC	ACTITACTIT	GCAGGGCTTC CCATAAAACC	CCAACCITAC	CAGAGGGGGG
3841	TGGCGGCGAG	AAAGCCAICC	CCTTGCTGT	CCATAAAACC	GCCCAGTCTA	GCTATCGCCA
3901	CCCAGCIGGC	AMITCUGGII	CCTGCTTTCT	CTTTGCGCTT	GCGTTTTCCC	TIGICCAGAI
3961	TGTAAGCCCA	CIGCHAGCIA	TCCGGGGTCA	GCACCGTTTC TABTCTCATG	TGCGGACTGG	CTTTCTACGT
4021	AGCCCAGTAG	CIGNONIA	TCCTTTTTGA	TAATCTCATG	ACCAAAATCC	CTTAACGTGA
4081	GANAAGGATC	CACCOLORAGE	CAGACCCCGT	AGAAAAGATC	AAAGGATCTT	CTTGAGATCC
4141	GTTTTCGTTC	CACTERATOTE	GCTGCTTGCA	AGAAAAAAA TTTTCCGAAG	CCACCGCTAC	CAGUGGIGGI
4201	TTTTTTTTT	CACCIARIOS	TACCAACTCT	TTTTCCGAAG CCCGTAGTTA	GTAACTGGCT	TCAGCAGAGC
4261	TIGITICCG	ANTACTGTCC	TTCTAGTGTA	GCCGTAGTTA AATCCTGTTA	GGCCACCACT.	TUARGARCIC
4321	GCAGATACCA	CCTACATACC	TEGETETGET	AATCCTGTTA AAGACGATAG	CCAGTGGCIG	CIGCCHGIGG
4381	TGTAGCACCG	TOTOTTACCG	GGTTGGACTC	AAGACGATAG CCCCAGCTTG	TTACCGGATA	AGGCGCAGCG
4441	CGATAAGICG	ACGGGGGGTT	CGTGCACACA	GCCCAGCTTG AAGCGCCACG	GAGCGAACGA	CCIACACCGA
4501	GICGGGCIGA	CTACAGCGTG	AGCTATGAGA	AAGCGCCACG AACAGGAGAG	CTTCCCGAAG	SCCTTCCAGG
4561	ACTOMORIAC	CCGGTAAGCG	GCAGGGTCGG	AACAGGAGAG CGGGTTTCGC	CCCACGAGGG	MICHECATOR
4621	GGACAGGIAI	TGGTATCTTT	ATAGTCCTGT	CGGGTTTCGC	CACCICIGAC	TIGACCICCTT
4681	PROGRAMACOCC	TECTCETCAG	GGGGGCGGAG	CCTATGGAAA TGCTCACATG	AACGCCAGCA	CCTTATCCCC
4741	WILLITIGICA	CTGGGCTTTT	GCTGGCCTTT	TGCTCACATG	TICTTICCIG	COLIMICOCO
4801						GCCGCAGGGG
4861 4921	TOWITCIGIO	CGCAGCGAGT	CAGTGAGCGA	GGAAGCGGAA	G	
44/1	TATO CATO CONTO				•	

2881

Figure 7

Composition 1522 A; 1620 C; 1590 G; 1577 T; 0 OTHER Percentage: 24% A; 26% C; 25% G; 25% T; 0%OTHER pMSVLSB4: 6309 bp;

dsDNA: 3889.6 Molecular Weight (kDa): ssDNA: 1947.08 AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC ORIGIN ACGACAGGIT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA 61 TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT 121 TTAGGTGACA CTATAGAATA CTCAAGCTAT GCATCAAGCT TGGTACCGAG CTCGGATCCA 181 CTAGTCCCGA TCTAGTAACA TAGATGACAC CGCGCGCGAT AATTTATCCT AGTTTGCGCG 241 CTATATITIG TTITCTATCG CGTATTAAAT GTATAATIGC GGGACTCTAA TCATAAAAAC 301 CCATCTCATA AATAACGTCA TGCATTACAT GTYAATTATT ACAYGCTEAA CGTAATTCAA 361 CAGAAATTAT. ATGATAATCA TCGACAGACC GGCAACAGGA TTCAATCTTA AGAAACTTTA 421 TTGCCAAATG TTTGAACGAT CGGGGAAATT CGCTCGAGTT AATTAAGCGG CCGCCTCAAA 481 AAGGATCITC ACCTAGATCC TITTAAATTA AAAATGAAGT TITAGCACGT GTCAGTCCTG 541 CTCCTCGGCC ACGAAGTGCA CGCAGTTGCC GGCCGGGTCG CGCAGGGCGA ACTCCCGCCC 601 CCACGGCTGC TCGCCGATCT CGGTCATGGC CGGCCCGGAG GCGTCCCGGA AGTTCGTGGA 661 CACGACCTCC GACCACTCGG CGTACAGCTC GTCCAGGCCG CGCACCCCACA CCCAGGCCAG 721. GGTGTTGTCC GGCACCACCT GGTCCTGGAC CGCGCTGATG AACAGGGTCA CGTCGTCCCG 781 GACCACACCG GCGAAGTCGT CCTCCACGAA GTCCCGGGAG AACCCGAGCC GGTCGGTCCA 841 GAACTCGACC GCTCCGGCGA CGTCGCGCGC GGTGAGCACC GGAACGGCAC TGGTCAACTT 901 GGCCATGGTG GCCCTCCTCA CGTGCTATTA TTGAAGCATT TATCAGGGTT ATTGTCTCAT 961 CAGCGGATAC ATATTTGAAT GTATTTAGAA AAATAAACAA ATAGGGGTTC CGCGCACATT 1021 TCCCCGAAAA GTGCCACCTG TATGCGGTGT GAAATACCGC ACAGATGCGT AAGGAGAAAA 1081 TACCGCATCA GGCGAAATTG TAAACGCGGC CGCTTAATTA AGTCGACGTC .CTCTCCAAAT 1141 GAAATGAACT TCCTTATATA GAGGAAGGGT CTTGCGAAGG ATAGTGGGAT TGTGCGTCAT 1201 CCCTTACGTC AGTGGAGATA TCACATCAAT CCACTTGCTT TGAAGACGTG GTTGGAACGT 1261 CITCITTTC CACGTAGCTC CTCGTGGGTG GGGGTCCATC TTTGGGACCA CTGTCGGCAG 1321 AGGCATCITG AACGATAGCC TITCCTTATC GCAATGATGG CATTIGTAGG TGCCACCITC 1381 CTTTTCTACT GTCCTTTGA TGAAGTGACA GATAGCTGGG CAATGGAATC CGAGGAGGTT 1441 TCCCGATATT ACCCTTTGTT GAAAAGTCTC AATAGCCCTT TGGTCTTCTG AGACTGTATC 1501 TITGATATIC TIGGAGTAGA CGAGAGAGTG TCGTGCTCCA CCATGTTGAC GAATTCATGG 1561 GCAGACCCGT CTGTACTTTA AGAGTGTTGG CAACCAGTAA TGAATAAAAA CTCCCGTTTT 1621 ATTATATTE ATGAATGCTG AAAGCTTACA TTAATATGTC GTGCGATGGC ACGAAAAAAC 1681 ACACGCAAAC AATACAGGGG GGTAGTCGGC GGGCGGCTAA GGGTGGTGCT CGGCGGGCAG 1741 AACATCGAAA AATCAAGATC TATATGAATT ACACTTCCTC CGTAGGAGGA AGCACAGGGG 1801 GAGAATACCA CTTCTCCCCC GGCGACATAA TGTAAATGAC GCAGTTTGCC TCGAAATACT 1861 CCAGCTGCCC TGGAGTCATT TCCTTCATCC AATCTTCATC CGAGTTGGCG AGGATTATTG 1921 TAGGCTTAGA CTTCTTCTGC ACCTTTTTCT TCTTACCATA CTTGGGGTTT ACAATGAAAT 1981 CCCTCTGACA GCCAACTAAC TGTTTCCAAC AAGGACAGAA TTTAAACGGA ATATCATCTA 2041 CGATGTTGTA GATTGCGTCT TCGTTGTATG AAGACCAATC AACATTATTT TGCCAGTAAT 2101 TATGAACCCC TAGGCTTCTG GCCCAAGTAG ATTTTCCGGT TCTTGTTGGG CCGACGATGT 2161 AGAGGCTCTG CTTTCTTGAT CTTTCATCTG ATGACTGGAT ACAGAATCCA TCCATTGGAG 2221 GTCAGAAATT GCATCCTCGA GGGTATAACA GGTAGGTTGA AGGAGGATGT AAGCTTCGGG 2281 ACTARCCTGG ARGATGTTAG GCTGGAGCCA ATCGTTGATT GACTCATTAC ARAGTARATC 2341 AGGTGAGGAG GGTGGATGAG GATTGGTGAA CTCTTCCTGA ATCTCAGGAA AAAGCTTATT 2401 TGCAGAGTAT TCAAAATACT GCAATTTTGT GGACCAATCA AAGGGGAGCT CTTTCTGGAT 2461 CATGGAGAGG TACTCTTCTT TGGAGGTAGC GTGTGAAATA ATGTCTCGCA TTATTTCATC 2521 TITAGAAGGC TITTITTCCT TTACCTCTGA ATCAGATTTT CCTAGGAAGG GGGACTTCCT 2581 AGGAATGAAA GTACCTCTCT CAAACACAGC CAGAGGTTCC TTGAGAATGT AATCCCTCAC 2641 TCTGTTAACT GACTTGGCAC TCTGAATATT TGGGTGAAAC CCATTTATAT CAAAGAACCT 2701 TGAGTCAGAT ATCCTTATCG GCTTCTCTGG CTGAAGCAAT GCATGTAAAT GCAAACTTCC 2761 ATCTTTATGT GCCTCTCGGG CACATAGAAT ATATTTGGGA ATCCAACGAA CGACGAGCTC 2821

Figure 7 (contd)

					mmente y CyC	TTTGGATAGG GAGGCCATAG	TTAGGAACGT
	2941 -	CCAGATCATC	TGACAGGCGA	TTTCAGGATT	CONTONCO	GAGGCCATAG CTATAGTATA	CCGACGACGG
	3001	GTTAGCGTTC	CTGTGTGAGA	ACTGACGGII	TOTAL A A CT	CTATAGTATA	CCCGTGCGCC
	3061	AGGTTGAGGC	TGAGGGATGG	CAGACIGGGA	OCTOR THE	TOGGCCGGAC	CGGGCCGGCC
	3121	TTCGAAATCC	GCCGCTCCAT	TGTCTTATAG	100110111	CHEMITICONGC	GAGGGCCCGG
	3181	CAGCAGGAAA	AGAAGGCGCG	CUCIANIAI	700000000	THE CALCULATION OF THE PARTY OF	TATCTAAAGC
	3241	GGTAGGGACC	GAGCGCTTTG	ATTTAMAGEC	1001122000	COTABCABGT	GCGATTCATT
	3301	AGCCCAATCT	AAAGAAACCG	GTCCCGGGCA	CINIMINA	COTATAGTGA	GTCGTATTAC
	3361	CATGGATCCT	TTAAACTCGA	GTCTAGAGGG		NCCCTCCCCCT	TACCCAACTT
	3421	AATTCACTGG	CCGTCGTTTT	ACAACGICGI	OVC10001211	DEDGCGAAGA	GGCCCGCACC
	3481	AATCGCCTTG	CAGCACATCC	CCCTTTCGCC	MY COMP CCCC	DOTTTAAGGT	TTACACCTAT
	3541	GATCGCCCTT	CCCAACAGTT	GCGCAGCCIA	21100211007A	CTCDTATTAT.	TGACACGCCG
	3601	AAAAGAGAGA	GCCGTTATCG	TCIGITIGIO	COLLO ZINCETCC	TOTOLGATAA	AGTCTCCCGT
	3661	GGGCGACGGA	TGGTGATCCC	CCIGGCCAGI	GAAAGCTGGC	GCATGATGAC ATCTCAGCCA	CACCGATATG
	3721	GAACTTTACC	CGGTGGTGCA	TATCGGGGAI	CARCTECTE	ATCTCAGCCA AAATGTCAGG	CCGCGAAAAT
	3781	GCCAGTGTGC	CGGTCTCCGT	TATCGGGGAA	TOCCOANTAT	AAATGTCAGG GTGGTGGTTA	CCTGAATGGC
	3841	GACATCAAAA	ACGCCATTAA	CCIGATGITC	CCCCCCCGGGT	GTGGTGGTTA GCTTTCTTCC	CGCGCAGCGT
	3901	GAATGGACGC	GCCCTGTAGC	GGCGCATIAA	OCCUPATION TO THE OCCUPATION OF THE OCCUPATION O	CCALLAL CALALCC	CTTCCTTTCT
,	3961	GACCGCTACA	CTTGCCAGCG	CCCTAGCGCC	COCTOCTATO	CCCCTCCCTT	TAGGGTTCCG
	4021	CGCCACGTTC	GCCGGCTTTC	CCCGTCAMGC	TOTAL TOTAL	THE CONTRACTOR	GTTCACGTAG
	4081	ATTTAGAGCT	TTACGGCACC	TCGACCGCAA	AND TOTAL	TTGGAGTCCA	CGTTCTTTAA
	4141	TGGGCCATCG	CCCTGATAGA	CGGTTTTTCG	ACTOBACCOT	TTGGAGTCCA ATCGCGGTCT AATGAGCTGA	ATTCTTTTGA
	4201	TAGTGGACTC	TIGITCCAAA	CIGGAACAAC	VCTCTATOOOT	ANTICACCTICA	ТТТААСАААА
	4261	TTTATAAGGG	ATGTTGCCGA	TTTCGGCCIA	CANGGOGATA	GAAGGCGATG CCATTCGCCG	CGCTGCGAAT
	4321	ATTITAACAA	AATTCAGAAG	AACTCGTCAA	MCCGGTCAGC	CCATTCGCCG GTCCGCCACA	CCAAGCTCTT
	4381	CGGGAGCGGC	GATACCGTAA	AGCACGAGGA	CTGATAGCG	GTCCGCCACA GATATTCGGC	CCCAGCCGGC
	4441	CAGCAATATC	ACGGGTAGCC	AACGCIAIGI.	TTTCCACCAT	GATATTCGGC CGCCTTGAGC	AAGCAGGCAT
	4501	CACAGTCGAT	GAATCCAGAA	MAGGGGGGGGT	CGGGCATGCT	CGCCTTGAGC ATCCTGATCG	CTGGCGAACA
	4561	CGCCATGGGT	CACGACGAGA	TCCTCGCCGT	CGTCCAGATC	ATCCTGATCG TTGGTGGTCG	ACAAGACCGG
	4621	GTTCGGCTGG	CGCGAGCCCC	TOWNSCION	GATGTTTCGC	TTGGTGGTCG CATGATGGAT	AATGGGCAGG
	4681	CTTCCATCCG	AGTACGTGCT	TCCCCCCCC	TTGCATCAGC	CATGATGGAT TTCGCCCAAT	ACTITCTCGG
	4741	TAGCCGGATC	AAGCGTATGU	ACCACATOTT	GCCCCGGCAC	TTCGCCCAAT AGGAACGCCC	AGCAGCCAGT
	4801	CAGGAGCAAG	GTGAGATGAC	ACCTCCAGCA	CAGCTGCGCA	AGGAACGCCC GGCACCGGAC	GTCGTGGCCA
	4861	CCCTTCCCGC	TTCAGTGACA	ACGICONOCI	GTTCATTCAG	GGCACCGGAC CACGGCGCA	AGGTCGGTCT
	4921	GCCACGATAG	CCCCCCTGCC	CCCTGCGCTG	ACAGCCGGAA	CACGGCGGCA CACCCAAGCG	TCAGAGCAGC
	4981	TGACAAAAAG	AACCGGGCGC	TOTAGCCGA	ATAGCCTCTC	CACCCAAGCG TCATCCTGTC	GCCGGAGAAC
	5041	CGATTGTCTG	TIGIGCCCAG	TCATACCOCC	GAAACGATCC	TCATCCTGTC AGCCATCCAG	TCTTGATCAG
	5101	CTGCGTGCAA	TCCATCTTGT	CAGATCCTTG	GCGGCGAGAA	AGCCATCCAG TTCCGGTTCG	TTTACTTTGC
	5161	ATCTTGATCC	CCTGCGCCAI	CAGACCCCCC	CAGCTGGCAA	TTCCGGTTCG GCAAGCTACC	CTTGCTGTCC
	5221	AGGGCTTCCC	AACCITACCA	TATTCCCCATG	TARGCCCACT	GCAAGCTACC GACATTCATC	TGCTTTCTCT
	5281	ATAAAACCGC	CCAGTCTAGC	CTCCAGATAG	CCCAGTAGCT	GACATTCATC GGTGAAGATC	CGGGGTCAGC
	5341	TTGCGCTTGC	GITTICCCII	GICCIONITIO	AAAGGATCTA	GGTGAAGATC CTGAGCGTCA	CTTTTTGATA
	5401	ACCGITICIG	CGGACTGGCT	TICINCOLOR T	TITCGTTCCA	CTGAGCGTCA CGTAATCTGC	GACCCCGTAG
	5461	ATCTCATGAC	CAAAATCCCT	TARCOTOTO	TITTTCTGCG	CGTAATCTGC TCAAGAGCTA	TGCTTGCAAA
	5521	AAAAGATCAA	AGGATCTTCT	CCCCCCCTTT	GTTTGCCGGA	TCAAGAGCTA TACTGTCCTT	CCAACTCTTT
	5581	CAAAAAAAACC	ACCGCTACCA	PCCPG1GG111	AGATACCAAA	TACTGTCCTT	CTAGTGTAGC
	5641	TTCCGAAGGT	AACTGGCTTC	AGCAGAGCOC	TAGCACCGCC	TACATACCTC TCTTACCGGG	GCTCTGCTAA
	5701	CGTAGTTAGG	CCACCACTIC	AAGAACICIO	ATABGTCGTG	TCTTACCGGG GGGGGGTTCG	TTGGACTCAA
	5761	TCCTGTTACC	AGTGGCTGCT	GCCHGIGGCO		CCCCCCTTCG	TGCACACAGC
	5821	GACGATAGTT	ACCGGATAAG	GCGCMGCGG1	TENGATACCT	ACAGCGTGAG	CTATGAGAAA
	5881	CCAGCTTGGA	GCGAACGACC	TACACCGARC	TOYOUTH TOO	CCTAACCGGC	AGGGTCGGAA
	5941	GCGCCACGCT	TCCCGAAGGG	AGAAAGGCGG	CALL DOUGHOUT COLUMN	СТАТСТТАТ	AGTCCTGTCG
	6001	CAGGAGAGCG	CACGAGGGAG	CTTCCAGGGG	CANTICOCO DE LA CONTRACTION DE	CTCGTCAGGG	GGGCGGAGCC
	6061	GGTTTCGCCA	CCTCTGACTT	GAGCGICGAI	TO COOKE COM	CCCCTTTTTGC	TGGCCTTTTG
	6121	TATGGAAAAA	CGCCAGCAAC	GCGGCCTTTT	THEORETECE	TARCCGTATT	ACCGCCTTTG
	6181	CTCACATGTT	CTTTCCTGCG	TTATCCCCIG	CCACCGAGCG	CAGCGAGTCA	GTGAGCGAGG
	6241	AGTGAGCTGA	TACCGCTCGC	CGCAGCCGAA	000000000000000000000000000000000000000	•	
	6301	AAGCGGAAG					

2881

Figure 8

pMSVLSB-5: 8043 bp; Composition 1983 A; 1992 C; 2011 G; 2057 T; 0 OTHER Percentage: 25% A; 25% C; 25% G; 26% T; 0%OTHER

Molecular Weight (kDa): ssDNA: 2483.31 dsDNA: 4958.5 AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC ORIGIN ACGACAGGIT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA 61 TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT 121 TTAGGTGACA CTATAGAATA CTCAAGCTAT GCATCAAGCT TGGTACCGAG CTCGGATCCA 181 CTAGTAACGG CCGCCAGTGT GCTGGAATTC ATGGGCAGAC CCGTCTGTAC TITAAGAGTG 241 TTGGCAACCA GTAATGAATA AAAACTCCCG TTTTATTATA TTTGATGAAT GCTGAAAGCT 301 TACATTARTA TGTCGTGCGA TGGCACGRAA AAACACACGC AAACAATACA GGGGGGTAGT 361 CGGCGGCGG CTAAGGGTGG TGCTCGGCGG GCAGAACATC GAAAAATCAA GATCTATATG 421 AATTACACTT CCTCCGTAGG AGGAAGCACA GGGGGAGAAT ACCACTTCTC CCCCGGCGAC 481 ATANTETAAN TGACGCAGIT TGCCTCGAAN TACTCCAGCT GCCCTGGAGT CATTTCCTTC 541 ATCCAATCIT CATCCGAGIT GGCGAGGATT ATTGTAGGCT TAGACTICIT CTGCACCITT 601 TTCTTCTTAC CATACTTGGG GTTTACAATG AAATCCCTCT GACAGCCAAC TAACTGFTTC 661 CAACAAGGAC AGAATTTAAA CGGAATATCA TCTACGATGT TGTAGATTGC GTCTTCGTTG 721 TATGAAGACC AATCAACATT ATTTTGCCAG TAATTATGAA CCCCTAGGCT TCTGGCCCAA 781 GTAGATITTC CGGTTCTTGT TGGGCCGACG ATGTAGAGGC TCTGCTTTCT TGATCTTTCA 841 TCTGATGACT GGATACAGAA TCCATCCATT GGAGGTCAGA AATTGCATCC TCGAGGGTAT 901 AACAGGTAGG TTGAAGGAGC ATGTAAGCTT CGGGACTAAC CTGGAAGATG TTAGGCTGGA 961 GCCAATCGTT GATTGACTCA TTACAAAGTA AATCAGGTGA GGAGGGTGGA TGAGGATTGG 1021 TGAACTCITC CTGAATCTCA GGAAAAGCT TATTTGCAGA GTATTCAAAA TACTGCAATT 1081 TTGTGGACCA ATCAAAGGGG AGCTCTTTCT GGATCATGGA GAGGTACTCT TCTTTGGAGG 1141 TAGCGTGTGA AATAATGTCT CGCATTATTT CATCTTTAGA AGGCTTTTTT TECTTTACCT 1201 CTGAATCAGA TTTTCCTAGG AAGGGGGGACT TCCTAGGAAT GAAAGTACCT CTCTCAAACA 1261 1381 CAGCCAGAGG TICCTIGAGA ATGTAATCCC TCACTCTGIT AACTGACTTG GCACTCTGAA TATTTGGGTG AAACCCATTT ATATCAAAGA ACCTTGAGTC AGATATCCTT ATCGGCTTCT CTGGCTGAAG CAATGCATGT AAATGCAAAC TTCCATCTTT ATGTGCCTCT CGGGCACATA 1441 GAATATATIT GGGAATCCAA CGAACGACGA GCTCCCAGAT CATCTGACAG GCGATTTCAG 1501 GATTITCTGG ACACTITGGA TAGGITAGGA ACGTGTTAGC GTTCCTGTGT GAGAACTGAC 1561 1681 GGTTGGATGA GGAGGAGGCC ATAGCCGACG ACGGAGGTTG AGGCTGAGGG ATGGCAGACT GGGAGCTCCA AACTCTATAG TATACCCGTG CGCCTTCGAA ATCCGCCGCT CCATTGTCTT ATAGTGGTTG TAAATGGGCC GGACCGGGCC GGCCCAGCAG GAAAAGAAGG CGCGCACTAA 1741 TATTACCGCG CCTTCTTTC CTGCGAGGGC CCGGTAGGGA CCGAGCGCTT TGATTTAAAG 1801 CCTGGTTCTG CTTTGTATGA TTTATCTAAA GCAGCCCAAT CTAAAGAAAC CGGTCCCGGG 1861 CACTATAAAT TGCCTAACAA GTGCGATTCA TTCATGGATC CTTTAAACTC GAGTCTAGTC 1921 CCGATCTAGT AACATAGATG ACACCGCGCG CGATAATTTA TCCTAGTTTG CGCGCTATAT 1981 TITGTTTTCT ATCGCGTATT ANATGTATAN TIGCGGGACT CTANTCATAN ANACCCATCT 2041 CATAAATAAC GTCATGCATT ACATGTTAAT TATTACATGC TTAACGTAAT TCAACAGAAA 2101 TTATATGATA ATCATCGACA GACCGGCAAC AGGATTCAAT CTTAAGAAAC TTTATTGCCA 2161 AATGTTTGAA CGATCGGGGA AATTCGCTCG AGTTAATTAA GCGGCCGCCT CAAAAAGGAT 2221 CTTCACCTAG ATCCTTTTAA ATTAAAAATG AAGTTTTAGC ACGTGTCAGT CCTGCTCCTC 2281 GGCCACGAAG TGCACGCAGT TGCCGGCCGG GTCGCGCAGG GCGAACTCCC GCCCCCACGG 2341 CTGCTCGCCG ATCTCGGTCA TGGCCGGCCC GGAGGCGTCC CGGAAGTTCG TGGACACGAC 2401 CTCCGACCAC TCGGCGTACA GCTCGTCCAG GCCGCGCACC CACACCCAGG CCAGGGTGTT 2461 GTCCGGCACC ACCTGGTCCT GGACCGCGCT GATGAACAGG GTCACGTCGT CCCGGACCAC 2521 ACCGGCGAAG TCGTCCTCCA CGAAGTCCCG GGAGAACCCG AGCCGGTCGG TCCAGAACTC 2581 GACCGCTCCG GCGACGTCGC GCGCGGTGAG CACCGGAACG GCACTGGTCA ACTTGGCCAT 2641 GGTGGCCCTC CTCACGTGCT ATTATTGAAG CATTTATCAG GGTTATTGTC TCATGAGCGG 2701 ATACATATTT GAATGTATTT AGAAAAATAA ACAAATAGGG GTTCCGCGCA CATTTCCCCG 2761 AAAAGTGCCA CCTGTATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC 2821

Figure 8 (cont'd)

		ATTGTAAACG		ATTANCTOGA	CCTCCTCTCC	AAATGAAATG
2941	ATCAGGCGAA	ATTGTAAACG TATAGAGGAA	CGGCCGCTIA	ALCATACTG	GGATTGTGCG	TCATCCCTTA
3001	AACTTCCTTA	TATAGAGGAA	GGGTCTTGCG	WAGGITTION -	CCTCCTTCGA	ACGTCTTCTT
3061	CGTCAGTGGA	GATATCACAT	CARICUACII	GC717GCCCC	NOCO CTCTCG	GCAGAGGCAT
3121	TTTCCACGTA	GCTCCTCGTG	GGTGGGGGTC	- magazina	TACCTCCCAC	CTTCCTTTTC
3181	CTTGAACGAT	AGCCTTTCCT	TATUGUARIO	MICCOLLEGE	NATOCCAGGA	GGTTTCCCGA
3241	TACTGTCCTT	TTGATGAAGT	GACAGATAGE		שרייונים בים נייונים	TATCTTTGAT
3301	TATTACCCTT	TGTTGAAAAG	TCTCAATAGC	TOTACONTOT	TCACCAATTC	ATGGGCAGAC
3361	ATTCTTGGAG	TAGACGAGAG	AGIGICUIGE	COLDENSTA	ANDOTTOTOG	TTTTATTATA
3421	CCGTCTGTAC	TTTAAGAGTG	TIGGCAACCA	011011011111	AGGCACCAAA	AAACACACGC
3481	TTTGATGAAT	GCTGAAAGCT	TACATIMALA		TOCTOGGGGG	GCAGAACATC
3541	AAACAATACA	GGGGGGTAGT	CGGCGGGCGG	CITATOCCATACC	NGGANGCACA	GGGGGAGAAT
3601	GAAAAATCAA	GATCTATATG	AATTACACTI	TOTAL COLOR COLOR	TOCCTOGAAA	TACTCCAGCT
3661	ACCACTTCTC	CCCCGGCGAC	ATARIGIAAA	TOTAL CONTRACTOR	CCCCACGATT	ATTGTAGGCT
3721	GCCCTGGAGT	CATTICCTIC	ATCCARICIT	CAT COURTER	CTTTACATG	AAATCCCTCT
3781	TAGACTICIT	CTGCACCTTT	TICITETIAC	AGAPTTTAÄÄ	CGGAATATCA	TCTACGATGT
3841	GACAGCCAAC	TAACTGTTTC	CAACAAGGAC	TT AD A ADDA	DATESTECAG	TAATTATGAA
3901	TGTAGATTGC	GTCTTCGTTG	TATGAAGACC	COMMONTON.	TYPECCUGACE	ATGTAGAGGC
3961	CCCCTAGGCT	TCTGGCCCAA	GTAGATTITC	COOTICION	ጥርርስጥርርስጣጥ	GGAGGTCAGA
4021	TCTGCTTTCT	TGATCTTTCA	TCIGATGACI	CONTROLL CONTROL	TODACTO	CGGGACTAAC
4081	AATTGCATCC	TCGAGGGTAT	AACAGGIAGG	TION COURS	TTECADAGTA	AATCAGGTGA
4141	CTGGAAGATG	TTAGGCTGGA	GCCAMICGII	Walter and a series	TODAGGGDS	TATTTGCAGA
4201	GGAGGGTGGA	TGAGGATTGG	TGAACICIIC	CIGARICICS	المالياليالياليال	GGATCATGGA
4261	GTATTCAAAA	TACTGCAATT	TTGTGGACCA	A TORNASOOO	CCATTATTT	CATCTTTAGA
4321	GAGGTACTCT	TCTTTGGAGG	TAGCGIGIGA	WILLIAM CO.	ANGGGGGACT	TCCTAGGAAT
4381	AGGCTTTTTT	TCCTTTACCT	CIGARICAGA		אייים מייבינים	TCACTCTGTT
4441	GAAAGTACCT	CTCTCAAACA	CAGCUAGAGG	1100110000	בשממתיים	ACCTIGAGIC
4501	AACTGACTTG	GCACTCTGAA	TATTIGGGIG	THE COURT	DAAAOSTAAAC	TTCCATCTTT
4561	AGATATCCTT	ATCGGCTTCT	CIGGCIGAAG		CONNCINCA	GCTCCCAGAT
4621	ATGTGCCTCT	CGGGCACATA	GAATATAT	- as ammodes	TACCTTAGGA	ACGTGTTAGC
4681	CATCTGACAG	GCGATITCAG	GATITICIO		ATAGCCGACG	ACGGAGGTTG
4741	GITCCIGIGI	GAGAACTGAC	GGTTGGATGA	cmcmama/	TATACCCCCCCCCC	CGCCTTCGAA
4801	AGGCTGAGGG	ATGGCAGACI'	GGGAGCICCA		CCNCCCCCCCC	GGCCCAGCAG
4861	ATCCGCCGCT	CCATTGTCTT	ATAGIGGIIO	The same of the sa	CTGCGAGGGC	CCGGGGTAGG .
4921	GAAAAGAAGG	CGCGCACTAA	TATTACCGCG	MCCOLOUR OUT IN	CATTTATTTA	AAGCAGCCCA
4981	GACCGAGCGC	TTTGATTTAA	AGCCIGGIAC	TOURS COMPANY	ANGRECCATT.	CATTCATGGA
5041	ATCTAAAGAA	ACCGGTCCCG	GGCACIAIAA	macadamama	CTCDCTCGTA	TTACAATTCA
5101	TCCTTTAAAC	TCGAGTCTAG	AGGGCCCAMI	100000000000000000000000000000000000000	CCCTTACCCA	ACTTAATCGC
5161	CTGGCCGTCG	TITTACAACG	TCGTGACTGG	OZERZANDO COG	AAGAGGCCCG	CACCGATCGC
5221	CTTGCAGCAC	ATCCCCCTTT	CGCCAGCIGG	CONTRACTOR	ACCITTACAC	CTATAAAAGA
5281	CCTTCCCAAC	AGTTGCGCAG	CCTATACGIA	CAGAGTGATA	TTATTGACAC	GCCGGGGCGA
5341	GAGAGCCGTT	ATCGTCTGTT	TGTGGATGIA	CTOTOTOTOTOTOTO	איים ממתית מיית	CCGTGAACTT
5401	CGGATGGTGA	TCCCCCTGGC	CAGTGCACGI	madecady area	TOACCACCGA	TATGGCCAGT
5461	TACCCGGTGG	TGCATATCGG	GGATGAAAGC	TOOCOCCATA	CCCACCGCGA	AAATGACATC
5521	GTGCCGGTCT	CCGTTATCGG	GGAAGAAGIG	ADADA A TOT	CAGGCCTGAA	TGGCGAATGG
5581	AAAAACGCCA	CCGTTATCGG TTAACCTGAT TAGCGGCGCA	GTTCTGGGGA	ATHIANATOL	GTTACGCGCA	GCGTGACCGC
5641	ACGCGCCCTG	TAGCGGCGCA	TTAAGCGCGC	GGGIGIGGIG	TTCCCCTTCCT	TTCTCGCCAC
5701	TACACTTGCC	TAGCGGCGCA AGCGCCCTAG	CCCCCCCTCC	TITUGUITE	COMMANGEER	TCCGATTTAG
5761	GTTCGCCGGC	AGCGCCCTAG TTTCCCCGTC	AAGCTCTAAA	TUGGGGGGCTC	CATTINGGOT	GTAGTGGGCC
5821	GTTCGCCGGC AGCTTTACGG	CACCTCGACC	GCAAAAAACT	TGATITGGGI	BCCyCcharcar	TTAATAGTGG
5881	ATCGCCCTGA	TAGACGGTTT	TTCGCCCTTT		עובל אובע עובאינים	TTGATTTATA
5941	ACTOTTGTTC	TAGACGGTTT CAAACTGGAA	CAACACTCAA	CCCTATUGUE	GICINITOTI	ATTTTAAAAA
6001	AGGGATGTTG	CAAACTGGAA CCGATTTCGG	CCTATTGGTT	AAAAAATGAG	CIGHTITHMC	GANTCGGGAG
6061	ACAAAATTCA	CCGATTTCGG GAAGAACTCG	TCAAGAAGGC	GATAGAAGGC	GATGCGCIGC	TCTTCAGCAA
6121	CGGCGATACC	GAAGAACTCG GTAAAGCACG	AGGAAGÒGGT	CAGCCCATTC	GUCGUCAAGU	CECCACAGT
6181	TATCACGGGT	GTAAAGCACG AGCCAACGCT	ATGTCCTGAT	AGCGGTCCGC	CACACCCAGC	GCATCGCCAT
6241	CGATGAATCC	AGCCAACGCT AGAAAAGCGG	CCATTTTCCA	CCATGATATT	CGGCAAGCAG	AACAGTTCGG
6301	GGGTCACGAC	AGAAAAGCGG GAGATCCTCG	CCGTCGGGCA	TECTCECCTT	GAGULTUGUU	AACAO11000
0201						

Figure 8 (cont'd)

			MALLICALICA A	GATCATCCTG	ATCGACAAGA	CCGGCTTCCA
6361	CTGGCGCGAG	CCCCTGATGC TGCTCGCTCG	TC11C01CC.	TOGOTTGGTG	GTCGAATGGG	CAGGTAGCCG
6421	TCCGAGTACG	TGCTCGCTCG ATGCAGCCGC	AIGCGAIGIA	CAGCCATGAT	GGATACTTTC	TCGGCAGGAG
6481	GATCAAGCGT	ATGCAGCCGC TGACAGGAGA	CGCATIGCAT	CCACTTCGCC	CAATAGCAGC	CAGTCCCTTC
6541	CAAGGTGAGA	TGACAGGAGA GACAACGTCG	1CCTGCCCCG	CCCAAGGAAC	GCCCGTCGTG	GCCAGCCACG
6601	CCGCTTCAGT	GACAACGTCG TGCCTCGTCT	AGCACAGCIG	TCAGGGCACC	GGACAGGTCG	GTCTTGACAA
6661	ATAGCCGCGC	TECCTCETCT	JOCHGIICKI	GGAACACGGC	GGCATCAGAG	CAGCCGATTG
6721	AAAGAACCGG	GCGCCCTGC CCAGTCATAG	GCTGACAGCC	TCTCCACCCA	AGCGGCCGGA	GAACCTGCGT
6781	TCTGTTGTGC	CCAGTCATAG	PROCESSE	ATCCTCATCC	TGTCTCTTGA	TCAGATCITG
6847	GCAATCCATC	TTGTTCAATC	ATGCGAMACG	AGAAAGCCAT	CCAGTTTACT	TTGCAGGGCT
6901	ATCCCCTGCG	CCATCAGATC ACCAGAGGGC	CITGGGGGGG	GCAATTCCGG	TTCGCTTGCT	GTCCATAAAA
6961	TCCCAACCIT	ACCAGAGGGC TAGCTATCGC	GCCCCAGC10	CACTGCAAGC	TACCTGCTTT	CTCTTTGCGC
7021						
7081	TIGCGITTIC	CCTTGTCCAG GGCTTTCTAC	MIMOCCOROL	TCTAGGTGAA	GATCCTTTTT	GATAATCTCA
7141	TCTGCGGACT	GGCTTTCTAC	CICHAMAGOR	TCCACTGAGC	GTCAGACCCC	GTAGAAAAGA
7201	TGACCAAAAT	CCCTTAACGT	CWGITITEGT	TGCGCGTAAT	CICCICCIIC	CAAACAAAAA
7261	TCAAAGGATC	TTCTTGAGAT ACCAGCGGTG	CCITITITIE	CGGATCAAGA	GCTACCAACT	CITTITECGA
7321						
7381						
7441	TAGGCCACCA	TGCTGCCAGT	TOTOTAGONE	COTOTCTTAC	CGGGTTGGAC	TCAAGACGAT
7501	TACCAGTGGC	TGCTGCCAGT TAAGGCGCAG	GGCGRIARGI	GAACGGGGG	TTCGTGCACA	CAGCCCAGCT
7561	AGTTACCGGA	TAAGGCGCAG GACCTACACC	COGICGOCCI	ACCTACAGOG	TGAGCTATGA	GAAAGCGCCA
7621	TGGAGCGAAC	GACCTACACC AGGGAGAAAG	GAACIGAGAS	ATCCGGTAAG	CGGCAGGGTC	GGAACAGGAG
7681	CGCTTCCCGA	AGGGAGAAAG GGAGCTTCCA	GCGGACAGG1	CCTGGTATCT	TTATAGTCCT	GTCGGGTTTC
7741	NGCGCACGAG	GGAGCTTCCA ACTTGAGCGT	GGGGGRRRCG	GATGCTCGTC	AGGGGGGGGG	AGCCTATGGA
7801	GCCACCTCTG	ACTTGAGCGT CAACGCGGCC	CONTILION	TCCTGGGCTT	TIGCIGGCCI	TTTGCTCACA
7861	AAAACGCCAG	CAACGCGGCC TGCGTTATCC	TITITIOGI	TGGATAACCG	TATTACCGCC	TTTGAGTGAG
7921	TGTTCTTTCC	TGCGTTATCC TCGCCGCAGC	CCIGHIICIG	AGCGCAGCGA	GTCAGTGAGC	GAGGAAGCGG
7981		TCGCCGCAGC	COMMICONCED		•	
8041	AAG	•	•			

2821

Figure 9

pMSVLSB-6: 7404 bp;

Composition 1839 A; 1794 C; 1835 G; 1936 T; 0 OTHER Percentage: 25% A; 24% C; 25% G; 26% T; 0%OTHER

dsDNA: 4564.5 Molecular Weight (kDa): ssDNA: 2286.33 AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC ORIGIN ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC 1 TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA 61 TIGIGAGCGG ATAACAATTI CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT 121 TTAGGTGACA CTATAGAATA CTCAAGCTAT GCATCAAGCT TGGTACCGAG CTCGGATCCA 181 CTAGTAACGG CCGCCAGTGT GCTGGAATTC ATGGGCAGAC CCGTCTGTAC TTTAAGAGTG 241 TTGGCAACCA GTAATGARTA AAAACTCCCG TTTTATTATA TITGATGAAT GCTGAAAGCT 301 TACATTAATA TGTGGTGCGA TGGCACGAAA AAACACACGC AAACAATACA GGGGGGTAGT 361 CGGCGGCGG CTAAGGGTGG TGCTCGGCGG GCAGAACATC GAAAAATCAA GATCTATATG 421 AATTACACTT CCTCCGTAGG AGGAAGCACA GGGGGAGAAT ACCACTTCTC CCCCGGCGAC 481 ATAATGTAAA TGACGCAGTT TGCCTCGAAA TACTCCAGCT GCCCTGGAGT CATTTCCTTC 541 ATCCAATCIT CATCCGAGIT GGCGAGGATT ATTGTAGGCT TAGACITCIT CTGCACCITT 601 TTCTTCTTAC CATACTTGGG GTTTACAATG AAATCCCTCT GACAGCCAAC TAACTGTTTC 661 CAACAAGGAC AGAATITAAA CGGAATATCA TCTACGATGT TGTAGATTGC GTCTTCGTTG 721 TATGANGACC ANTCANCATT ATTITECCAG TANTIATGAN CCCCTAGGCT TCTGGCCCAN 781 GTAGATTTTC CGGTTCTTGT TGGGCCGACG ATGTAGAGGC TCTGCTTTCT TGATCTTTCA 841 TCTGATGACT GGATACAGAA TCCATCCATT GGAGGTCAGA AATTGCATCC TCGAGGGTAT 901 AACAGGTAGG TTGAAGGAGC ATGTAAGCTT CGGGACTAAC CTGGAAGATG TTAGGCTGGA 951 GCCAATCGTT GATTGACTCA TTACAAAGTA AATCAGGTGA GGAGGGTGGA TGAGGATTGG 1021 TGAACTCITC CTGAATCTCA GGAAAAAGCT TATTTGCAGA GTATTCAAAA TACTGCAATT 1081 TTGTGGACCA ATCAAAGGGG AGCTCTTTCT GGATCATGGA GAGGTACTCT TCTTTGGAGG 1141 TAGCGTGTGA ANTANTGTCT CGCATTATIT CATCTITAGA AGGCTTTTTT TCCTTTACCT 1201 CTGAATCAGA TTTTCCTAGG AAGGGGGACT TCCTAGGAAT GAAAGTACCT CTCTCAAACA 1261 CAGCCAGAGG TTCCTTGAGA ATGTAATCCC TCACTCTGTT AACTGACTTG GCACTCTGAA 1321 TATITEGGTG ANACCCATTT ATATCANAGA ACCTTGAGTC AGATATCCTT ATCGGCTTCT 1381 CTGGCTGAAG CAATGCATGT AAATGCAAAC TTCCATCTTT ATGTGCCTCT CGGGCACATA 1441 GAATATATIT GGGAATCCAA CGAACGACGA GCTCCCAGAT CATCTGACAG GCGATTTCAG 1501 GATTITCTGG ACACTITGGA TAGGTTAGGA ACGTGTTAGC GTTCCTGTGT GAGAACTGAC 1561 GGTTGGATGA GGAGGAGGCC ATAGCCGACG ACGGAGGTTG AGGCTGAGGG ATGGCAGACT 1621 GGGAGCTCCA AACTCTATAG TATACCCGTG CGCCTTCGAA ATCCGCCGCT CCATTGTCTT 1681 ATACTCCTTC TARATCCCCC GCACCCGCCC GCCCCAGCAG GAAAAGAAGC CCCCCACTAA 1741 TATTACCECE CCTTCTTTC CTGCGAGGGC CCGGTAGGGA CCGAGCGCTT TGATTTAAAG 1801 CCTGGTTCTG CTTTGTATGA TTTATCTAAA GCAGCCCAAT CTAAAGAAAC CGGTCCCGGG 1861 CACTATAAAT TGCCTAACAA GTGCGATTCA TTCATGGATC CTTTAAACTC GAGTCTAGTC 1921 CCGATCTAGT AACATAGATG ACACCGCGCG CGATAATTTA TCCTAGTTTG CGCGCTATAT 1981 TITGITITCT ATCGCGTATT AAATGTATAA TIGCGGGACT CTAATCATAA AAACCCATCT 2041 CATAAATAAC GTCATGCATT ACATGTTAAT TATTACATGC TTAACGTAAT TCAACAGAAA 2101 TTATATGATA ATCATCGACA GACCGGCAAC AGGATTCAAT CTTAAGAAAC TTTATTGCCA 2161 AATGTTTGAA CGATCGGGGA AATTCGCTCG AGTTAATTAA GCGGCCGCTT AATTAAGTCG 2221 ACGTCCTCTC CARATGARAT GARCTTCCTT ATATAGAGGA AGGGTCTTGC GAAGGATAGT 2281 GGGATTGTGC GTCATCCCTT ACGTCAGTGG AGATATCACA TCAATCCACT TGCTTTGAAG 2341 ACGTGGTTGG AACGTCTTCT TTTTCCACGT AGCTCCTCGT GGGTGGGGGT CCATCTTTGG 2401 GACCACTGTC GGCAGAGGCA TCTTGAACGA TAGCCTTTCC TTATCGCAAT GATGGCATTT 2461 GTAGGTGCCA CCTTCCTTTT CTACTGTCCT TITGATGAAG TGACAGATAG CTGGGCAATG 2521 GAATCCGAGG AGGTITCCCG ATATTACCCT TTGTTGAAAA GTCTCAATAG CCCTTTGGTC 2581 TTCTGAGACT GTATCTTGA TATTCTTGGA GTAGACGAGA GAGTGTCGTG CTCCACCATG 2641 TTGACGAATT CATGGGCAGA CCCGTCTGTA CTTTAAGAGT GTTGGCAACC AGTAATGAAT 2701 ARARACTCCC GTTTTATTAT ATTTGATGAR TGCTGARAGC TTACATTART ATGTCGTGCG 2761

Figure 9 (cont'd)

					·	•
2881	ATGGCACG	AA AAAACACA	CG CAAACAAT	AC AGGGGGGTA	G TCGGCGGGCG	GCTAAGGGTG
2941	GTGCTCGG	CG GGCAGAAC	AT CGAAAAAT	CA AGATCTATA	T GAATTACACI	TCCTCCGTAG
3001	GAGGAAGC	AC AGGGGGAG	AA TACCACTT	CT CCCCCGGCG	A CATAATGTAA	ATGACGCAGT
3061	TTGCCTCG	AA ATACTCCA	GC TGCCCTGG	AG TCATTTCCT	r CATCCAATCI	TCATCCGAGT
3121	TGGCGAGG	AT TATTGTAG	GC TTAGACTT	CT TCTGCACCT	r TITCITCITA	CCATACTIGG
3181	GGTTTACA	AT GAAATCCC	TC TGACAGCC	AA CTAACTGTT	r ccaacaagga	CAGAATTTAA
3241	ACGGAATA'	rć atctacga	TG TTGTAGAT	IG CGTCTTCGT	r GTATGAAGAC	CAATCAACAT
3301	TATTTTGC	CA GTAATTAT	ga acccctage	C TTCTGGCCC	AGTAGATIT	CCGCTTCTTG
3361	TTGGGCCG	AC GATGTAGA	GG CTCTGCTT	C TIGATCITIC	ATCTGATGAC	TGGATACAGA
3421	ATCCATCC	AT TGGAGGTC	AG AAATTGCAT	CTCGAGGGT	TAACAGGTAG	GTIGAAGGAG
3481	CATGTAAG	T TCGGGACT	aa cctggaagi	AT GTTAGGCTGG	AGCCAATCGT	TGATTGACTC
3541	ATTACAAAC	T AAATCAGG	IG AGGAGGGT	G ATGAGGATTO	GTGAACTCIT	CCIGAATCIC
3601	aggaaaaa	C TTATTTGC	AG AGTATTCAP	A ATACTGCAAT	TTTGTGGACC	AATCAAAGGG
3661	GAGCTCTTT	C TGGATCAT	eg agaggtaci	C TTCTTTGGAG	GTAGCGTGTG	AAATAATGTC
3721	TCGCATTAT	T TCATCTTI	AG AAGGCTTTI	T TICCITIACO	TCTGAATCAG	ATTITCCTAG
3781	GAAGGGGGA	C TTCCTAGG	AN TGAŅAGTAC	C TCTCTCAAAC	ACAGCCAGAG	GITCCITGAG
3841	AATGTAATC	C CTCACTCTY	TAACTGACT	T GGCACTCTGA	ATATTTGGGT	GAAACCCATT
3901	TATATCAAA	G AACCTTGAG	T CAGATATCC	T TATCGGCTTC	TCTGGCTGAA	GCAATGCATG
3961	TAAATGCAA	A CITCCATCI	T TATGTGCCT	C TEGGGEACAT	AGAATATATT	TGGGAATCCA
4021	ACGAACGAC	G AGCTCCCAG	A TCATCTGAC	A GGCGATITCA	GGATTTTCTG	GACACITIGG
4081	ataggttag	G AACGTGTTA	G CGITCCIGI	G TGAGAACTGA	CCGTTGGATG	AGGAGGAGGC
4141	CATAGCCGA	C GACGGAGGI	T GAGGCTGAG	G GATGGCAGAC	TGGGAGCTCC	AAACTCTATA
4201	GTATACCCG	T GCGCCTTCG	A AATCCGCCG	C TCCATTGTCT	TATAGTGGTT	GTAAATGGGC
4261	CGGACCGGG	C CGGCCCAGC	A GGAAAAGAA	G GCGCGCACTA	ATATTACCGC	GCCTTCTTTT
4321	CCTGCGAGG	G CCCGGGGTA	G GGACCGAGC	G CTTTGATTTA	AAGCCTGGTT	CIGCITIGIA
.4381	TGATTTATC	r aaagcagcc	C AATCTAAAG	A AACCGGTCCC	GGGCACTATA	MATIGUCIAA
4441	CAAGTGCGA:	I TCATTCATG	G ATCCTTTAA	A CTCGAGTCTA	CAGGGGCCCAA	TICGCCCIMI
4501	AGTGAGTCG:	r ATTACAATT	C ACTGGCCGT	CATCCCCCTT	TOCOCOLORCIO	CCTAATACC
4561	GGCGTTACCC	AACTTAATO	G CCTTGCAGCA	A CAGTIGCGCA	CCCTATACIO	DCCLARLINGO DCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
4621	GAAGAGGCC	GCACCGATO	i resestedes	TATCGTCTGT	THEREGARET	CAGAGTGAT
4681	AAGGTTTACA	CCIAIAAAA	a ACCCATCCTC	ATCCCCCTGG	CCAGTGCACG '	TCTGCTGTCA
4741	ATTATTGACA		P PERCECTOR	GTGCATATCG	GGGATGAAAG	TGGCGCATG
4801	GATAAAGICI >maxaaa	L CCCGIGGAC	. 4545CCC661C	TCCGTTATCG	GGGAAGAAGT (GCTGATCTC
4861	ATGACCACCO	AIRIGGCUR		ATTAACCTGA	TGTTCTGGGG A	DTATATAATG
4921	TCA CCCCTCA	, WHUMIGHOR	CARACACCC	GTAGCGGCGC	ATTANGCGCG (CCCTCTCCT
4981 5041	CCAGGCCIGA	NOCOTORIO	CTACACTTGC	CAGCGCCCTA	GCGCCCGCTC (TTTCGCTTT
5101	CTTCCCTTCC	. 2000310200	CGTTCGCCGG	CTTTCCCCGT	CAAGCTCTAA A	TCGGGGGCT
5161	CIICCCIICC	יייייייייייייייייייייייייייייייייייייי	CACCTTTACG	GCACCTCGAC	CGCAAAAAAC T	TGATTTGGG
5221	TOTAL TANGE	CONTRACTOR	CATCGCCCTG	ATAGACGGTT	TITCGCCCTT I	GACGTTGGA
5281	TOTOTOTO	TTTAATAGTG	GACTCTTGTT	CCAAACTGGA	ACAACACTCA A	CCCTATCGC
5341	CCTCTATTCT	מתידים מבצייים י	AAGGGATGTT	GCCGATTTCG	GCCTATTGGT T	AAAAAAY
5401	GCTCINTICI	דדידעמממי	PACABABATTC	AGAAGAACTC	TCAAGAAGG C	GATAGAAGG
5461	GCIGNITIAN		GCGCCGATAC	CGTAAAGCAC	SAGGAAGCGG T	CAGCCCATT
	COAIGCGCIG	CONTICOCO	ATATCACGG	TAGCCAACGC	PATGTCCTGA T	AGCGGTCCG
5521	CGCCGCCAAG	CICIICAGUA	TOTATEDATE	CAGAAAAGCG (CCATTTTCC A	CCATGATAT
5581 5643	TOTAL ACCUAG	CCGGCCACAG	TOGGTCACGA	CGAGATCCTC (CCGTCGGGC A	TGCTCGCCT
5641	TCGGCAAGCA	CANCACTUCC	CCTCCCCCC	GCCCTGATG	TCTTCGTCC A	GATCATCCT
5701	TGAGCCTGGC	DAACAGIICG	AUCCCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	GTGCTCGCTC (CATCOCCATCT T	TCGCTTGGT
5761	GATCGACAAG	ACCEGETICE	CCATCAACCG	TATGCAGCCG (CGCATTGCA T	CAGCCATGA
5821	OCTUGARIOU	GCHOOTHOCC	GCATCARGO	ATGACAGGAG A	TCCTGCCCC G	GCACTTCGC
5881	1GGATACTIT	CICOCCAGGA	CONGRETANG	TGACAACGTC	BAGCACAGCT G	CGCAAGGAA
5941	CCAATAGCAG	CCAGTACCTT	COCOCITORS	CTGCCTCGTC 1	TGCAGTTCA T	TCAGGGCAC
6001 6061	CCCCCGTCGT	CONCORPORA	DATABLCOCO	GCCCCCTG C	GCTGACAGC O	GGAACACGG
6061	COCACAGGIC	CONCCOUNT	WHANGHHOUGH WHANGHHOUGH	CCCAGTCATA G	CCGAATAGC C	TCTCCACCC
5121 5101	ANGCCCCCCCC	DEADCCOATT	GICIGIIGIO	CTTGTTCAAT C	ATGCGAAAC G	ATCCTCATC
5181 5241	MAGCGGCCGG	MONNECTOCK	TOCHWICCHI	GCCATCAGAT C	CTTGGCGGC G	AGAAAGCCA
ノチュエ	CIGICICIIG	WICHOUT CTT				-

Figure 9 (cont'd)

				ma contractor	CCCCCACCT	GGCAATTCCG
6301	TCCAGTTTAC	TTTGCAGGGC	TTCCCAACCT	TACCAGAGGG	CGCCCCAGCT	CCACTICCAAG
6361			\sim	CTAGCTATUG	CCMIGITAGE	CONCACCARO
			بالململيات المتعلف	CCCTTGTCCA	CATACCCCAG	TWOCTOWCHT
6421			DAISONSWIP TO THE PROPERTY OF	TGGCTTTUTA	CGIGHMMAGG	VICTUOGICH
6481			- ~~~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~	TITITAALG	IGNOTITIO	
6541	AGATCCITTT	TGATAATCIC	AIGACGEE.	CHICHTOTAGA	TCCTTTTTT	CTGCGCGTAA
6601	CGTCAGACCC	CGTAGAAAAG	ATCAAAGGAI	ma coaccer	CCTTTCTTTG	CCGGATCAAG
6661	TCTGCTGCTT	GCAAACAAAA	ANACCACCGC	TACCAGCGGI	GGTTTGTTTG	בעים מידע מענים
6721			እ እርፈጥስ እርጣፕ	GCTTCAGCAG	MOCOUNTRY	
6781				ACTITIC AAGAA	CICIGIAGOS	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
•				CALL SELL SELL	TURELLERING	TOTOTOTIV
6841		~~~~~~~~	AND CALLED VALE	ATAMETER	000010000	
6901						
6961	GTTCGTGCAC	ACAGCCCAGC	TIGGAGCGAA	PAGGGAGAAA	GGCGGACAGG	TATCCGGTAA
7021	GTGAGCTATG	AGAAAGCGCC.	ACGCTTCCCG	MAGGGAGAAA	POCCESTATE	CCCTGGTATC
7081		~~~~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~	CACCCCACTIA	(ACHIDINAL LA L	WOOOOO WAYNA	
7141			~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	CADI "I"ITTANITALIS	ICGWITTING	2022202
		CO COOMS EVOVS	AAAAAACCCCA	GCAACGCGG	CITITIANCO	******
7201			Market All In In 1 and I all I a	CIGCGITUTE	CCCICALA	
7261			TOTACTAC	CHARCAGCAG	CCGAACGACC	GAGCGCAGCG
7321				C10000000		
7201	AGTCAGTGAG	CGAGGAAGCG	GAAG			•

SEQUENCE LISTING

•	<110> LARGE SCALE BIOLOGY CORPORATION	
,	<120> COMPOSITIONS AND METHODS FOR INHIBITING GENE EXPRESSION	
•	<130> 008010177PC00	
	<140> To Be Assigned	
	<141> 2001-04-04	
	<150> 09/545,574	
	<151> 2000-04-07	
•	<160> 14	
	<170> FastSEQ for Windows Version 3.0	
	<210> 1	
	<211> 27	
	<212> DNA	
	<213> Cauliflower mosaic virus	
	<400> 1	27
tttgaa	ttcg tcaacatggt ggagcac	. /
	<210> 2	
	<211> 31	
	<211> DNA	
	<213> Cauliflower mosaic virus	
	<400> 2	31
tttgtc	gacg teeteteeaa atgaaatgaa e) I
	<210> 3	
	<211> 46	
	<211> 40 <212> DNA	
	<213> Artificial Sequence	
	<220>	
	<223> zeocin resistance gene	
	<400> 3	46
cccgtc	gact taattaageg geegegttta caatttegee tgatge	
	<210> 4	
	<211> 47	
	<212> DNA	
	<213> Artificial Sequence	
•	- ,	
	<220>	
	<223> zeocin resistance gene	
	400 4	
	<400> 4 gagt taattaageg geegeeteaa aaaggatett cacetag	4
	udul ladilaayiy yilyiciida aaayyarore uuuuus	

<210> 5	
<211> 32	
<212> DNA	
<213> Artificial Sequence	
•	
<220>	
<223> nopaline synthase gene (nos) terminator sequence	
<400> 5	
tttctcgagc gaatttcccc gatcgttcaa ac	32
<210> 6	•
<211> 32	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> nopauline synthase (nos) terminator sequence	
<400> 6	20
tttactagtc ccgatctagt aacatagatg ac	32
<211> 29	
<212> DNA	
<213> maize	
<400> 7	29
tttttaatta aggtccgcct gaattctcg	
<210> 8	
<211> 30	
<211> 30 <212> DNA	
<213> maize	
(213) IIIa126	
<400> 8	
tttttaatta acggcaaggc tcacagtttg	30
coccaacca acggeatgge coacagootg	
<210> 9	
<211> 4881	
<212> DNA	
<213> Viral	
<400> 9	
agegeceaat aegeaaaceg eeteteeeeg egegttggee gatteattaa tgeagetgge	60
acqacaggtt tecegaetgg aaagegggea gtgagegeaa egeaattaat gtgagttage	120
tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa	180
ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac gccaagctat	240
ttaggtgaca ctatagaata ctcaagctat gcatcaagct tggtaccgag ctcggatcca	300
ctagtaacgg ccgccagtgt gctggaattc atgggcagac ccgtctgtac tttaagagtg	360
ttggcaacca gtaatgaata aaaactcccg ttttattata tttgatgaat gctgaaagct	420
tacattaata tgtcgtgcga tggcacgaaa aaacacacgc aaacaataca ggggggtagt	480
cggcgggcgg ctaagggtgg tgctcggcgg gcagaacatc gaaaaatcaa gatctatatg	540
aattacactt cctccqtagg aggaagcaca gggggagaat accacttctc ccccggcgac	600
ataatgtaaa tgacgcagtt tgcctcgaaa tactccagct gccctggagt catttccttc	- 660
	720

	ttcttcttac	catacttggg	gtttacaatg	aaatccctct	gacagccaac	taactgtttc	780
	caacaaqqac	agaatttaaa	cggaatatca	tctacgatgt	tgtagattgc	gtcttcgttg	840
	tatgaagacc	aatcaacatt	attttgccag	taattatgaa	cccctaggct	tctggcccaa	900
	gtagattttc	cggttcttgt	tgggccgacg	atgtagaggc	tetgetttet	tgatcttca	960
	totgatğact	ggatacagaa	tccatccatt	ggaggtcaga	aattgcatcc	tcgagggtat	1020
	aacaqqtagg	ttgaaggagc	atgtaagctt	cgggactaac	ctggaagatg	ttaggctgga	1080
	gccaatcgtt	gattgactca	ttacaaagta	aatcaggtga	ggagggtgga	tgaggattgg	1140
	tgaactcttc	ctgaatctca	ggaaaaagct	tatttgcaga	gtattcaaaa	tactgcaatt	-1200
	ttqtqqacca	atcaaaqqqq	agctctttct	ggatcatgga	gaggtactct	tctttggagg	1260
	tagogtgtga	aataatqtct	cgcattattt	catctttaga	aggctttttt	tcctttacct	1320
	ctgaatcaga	ttttcctagg	aaqqqqqact	tcctaggaat	gaaagtacct	ctctcaaaca	1380
	cagccagagg	ttccttgaga	atgtaatccc	tcactctgtt	aactgacttg	gcactctgaa	1440
	tatttaggta	aaacccattt	atatcaaaga	accttgagtc	agatatcctt	atcggcttct	1500
	ctggctgaag	caatgcatgt	aaatqcaaac	ttccatcttt	atgtgcctct	cgggcacata	1560
	gaatatattt	gggaatccaa	cqaacqacqa	gctcccagat	catctgacag	gcgatttcag	1620
	gattttctgg	acactttqqa	taggttagga	acgtgttagc	gttcctgtgt	gagaactgac	1680
	gattagatga	ggaggaggcc	atagecgaeg	acqqaqqttq	aggctgaggg	atggcagact	1740
	gggagctcca	aactctataq	tatacccqtq	caccttcqaa	atccgccgct	ccattgtctt	1800
	atagtggttg	taaatgggcc	ggaccgggcc	ggcccagcag	gaaaagaagg	cgcgcactaa	1860
	tattaccoco	ccttcttttc	ctacaàaaac	ccaataaaaa	ccgagcgctt	tgatttaaag	1920
	cctaattcta	ctttacaacc	actcgagcat	gcatctagag	ggcccaattc	gccctatagt	1980
	gagrogtatt	acaattcact	gaccatcatt	ttacaacqtc	gtgactggga	aaaccctggc	2040
	gttacccaac	ttaatcgcct	tgcagcacat	cccctttcq	ccagctggcg	taatagcgaa	2100
	gecacccaca	ccgatcgccc	ttcccaacaq	ttacacaacc	tatacgtacg	gcagtttaag	2160
	green	ataaaagaga	gagccottat	catctattta	tggatgtaca	gagtgatatt	2220
	attaacacac	caaaacasca	gatogtgatc	cccctggcca	gtgcacgtct	gctgtcagat	2280
	assorctccc	grgaacttta	cccaataata	catatoggg	atgaaagctg	gcgcatgatg	2340
	accaccgata	taaccaatat	accaatetee	gttatcgggg	aagaagtggc	tgatctcagc	2400
	caccaccaca	atgacatcaa	aaacgccatt	aacctgatgt	tctggggaat	ataaatqtca	2460
	acctasta	acquatagac	gcgccctgta	gcggcgcatt	aagcgcgcgg	gtgtggt	2520
	tacacacaac	graaccacta	cacttoccag	caccctaaca	cccgctcctt	tcgctttctt	2580
	cccttccttt	ctcaccacat	tcaccaactt	tccccqtcaa	gctctaaatc	gggggctccc	2640
	tttagggttc	coatttagag	ctttacggca	cctcgaccgc	aaaaaacttg	atttgggtga	2700
	taattaacat	agtoggccat	caccctgata	gacggttttt	cgccctttga	cattagaatc	2760
	cacattatt	aatagtggac	tettetteca	aactggaaca	acactcaacc	ctatcgcggt	2820
	ctattcttt	gatttataag	ggatgttgcc	gatttcggcc	tattggttaa	aaaatqaqct	2880
	catttaacaa	aaattttaac	aaaattcaga	agaactcgtc	aagaaggcga	taqaaggcga	2940
	tacactacaa	atcoppage	acgataccat	aaagcacgag	gaagcggtca	gcccattcgc	3000
•	caccaaactc	ttcagcaata	tracqqqtaq	ccaacqctat	gtcctgatag	caatccacca	3060
	caccaaacca	accacaatca	atgaatccag	aaaagcggcc	attttccacc	atgatattcg	3120
	casecages	atcaccatag	atcacgacga	gatectegee	gtcgggcatg	ctcqccttqa	3180
	acctaacaa	cagttcggct	aacacaaacc	cctgatgctc	ttcgtccaga	tcatcctgat	3240
	descalace	agettecate	caagtacata	ctcactcaat	gcgatgtttc	acttaataat	3300
	cgacaagace	agtageega	tcaagcgtat	acaaccacca	cattgcatca	qccatqatqq	3360
	atactttctc	ggcagcegga	aggtgagatg	acaggagatc	ctgccccggc	acttcqccca	3420
	ataggagga	gtcccttccc	gcttcagtga	caacgtcgag	cacagetgeg	caaqqaacqc	3480
	ccatcataac	cagccacgat	agccacacta	cctcatctta	cagttcattc	agggcaccgg	3540
•	acadatogat	cttgacaaaa	agaaccgggC	accetacae	tgacagccgg	aacacqqcqq	3600
	catcagagge	accastrate	tattatacce	agtcatagee	gaatagcctc	tccacccaaq	3660
	caaccaasas	acctacatac	aatccatctt	gttcaatcat	gcgaaacgat	cctcatcctq	3720
	tetetteate	agatetteat	ccctacacc	atcagatect	tggcggcgag	aaagccatcc	3780
	agtttagttt	gcagggette	ccaaccttac	cagaggggg	cccagctggc	aattccggtt	3840
	cacttactat	ccataaaacc	acceatora	gctatcgcca	tgtaagccca	ctgcaagcta	3900
	cctactttct	ctttacactt	acatttteee	ttatccagat	agcccagtag	ctgacattca	3960
	tecagactes	gcaccottte	tacaaactaa	ctttctacot	gaaaaggatc	taggtqaaqa	4020
	teetttte	taatctcato	accassator	cttaacotoa	gttttcgttc	cactgagcgt	4080
	cadacccccat	agaaaagate	aaaggatett	cttgagatcc	tttttttctg	cacataatct	4140
	Lagactety	agadaagatt	~~~			J : J :	

PCT/US01/11436 WO 01/77350

```
gctgcttgca aacaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc
                                                                      4200
                                                                      4260
taccaactet tttteegaag gtaactgget teageagage geagatacea aatactgtee
                                                                      4320
ttctagtgta gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc
                                                                      4380
togototgot aatootgtta coagtggotg otgocagtgg ogataagtog tgtottacog
                                                                      4440
ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga acggggggtt
                                                                      4500
cgtgcacaca gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg
agctatgaga aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg
                                                                      4560
                                                                      4620
gcagggtcgg aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggtatcttt
atagtectgt egggtttege cacetetgae ttgagegteg atttttgtga tgetegteag
                                                                      4680
gggggcggag cctatggaaa aacgccagca acgcggcctt tttacggttc ctgggctttt
                                                                      4740
gctggccttt tgctcacatg ttctttcctg cgttatcccc tgattctgtg gataaccgta
                                                                      4800
                                                                      4860
ttaccgcctt tgagtgagct gataccgctc gccgcagccg aacgaccgag cgcagcgagt
                                                                      4881
cagtgagcga ggaagcggaa g
      <210> 10
      <211> 3413
      <212> DNA
      <213> viral
      <400> 10
                                                                        60
agegeceaat aegeaaaceg ceteteceeg egegttggee gatteattaa tgeagetgge
acgacaggtt tecegactgg aaagegggea gtgagegeaa egeaattaat gtgagttage
                                                                       120
tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa
                                                                       180
                                                                       240
ttgtgagegg ataacaattt cacacaggaa acagetatga ecatgattae gecaagetat
ttaggtgaca ctatagaata ctcaagctat gcatcaagct tgggcccggt agggaccgag
                                                                       300
                                                                       360
egetttgatt taaageetgg ttetgetttg tatgatttat etaaageage eeaatetaaa
gaaaccggtc ccgggcacta taaattgcct aacaagtgcg attcattcat ggatccttta
                                                                       420
                                                                        480
aactcgagtc tagagggccc gaattctgca gatatccatc acactggcgg ccgctcgagc
                                                                        540
atgcatctag agggeccaat tegecetata gtgagtegta ttacaattea etggeegteg
                                                                        600
ttttacaacg tcgtgactgg gaaaaccctg gcgttaccca acttaatcgc cttgcagcac
atcccccttt cgccagctgg cgtaatagcg aagaggcccg caccgatcgc ccttcccaac
                                                                        660
                                                                        720
agttgcgcag cctatacgta cggcagttta aggtttacac ctataaaaga gagagccgtt
                                                                        780
atcgtctgtt tgtggatgta cagagtgata ttattgacac gccggggcga cggatggtga
                                                                        840
tececetgge cagtgeacgt etgetgteag ataaagtete eegtgaactt taeeeggtgg
                                                                        900
tgcatatcgg ggatgaaagc tggcgcatga tgaccaccga tatggccagt gtgccggtct
                                                                        960
ccgttatcgg ggaagaagtg gctgatctca gccaccgcga aaatgacatc aaaaacgcca
                                                                       1020
ttaacctgat gttctgggga atataaatgt caggcctgaa tggcgaatgg acgcgccctg
                                                                      1080
tageggegea ttaagegege gggtgtggtg gttaegegea gegtgaeege taeaettgee
agegeectag egecegetee tttegettte tteeetteet ttetegeeae gttegeegge
                                                                       1140
tttccccgtc aagctctaaa tcgggggctc cctttagggt tccgatttag agctttacgg
                                                                       1200
cacctcgacc gcaaaaaact tgatttgggt gatggttcac gtagtgggcc atcgccctga
                                                                       1260
tagacggttt ttcgcccttt gacgttggag tccacgttct ttaatagtgg actcttgttc
                                                                       1320
caaactggaa caacactcaa ccctatcgcg gtctattctt ttgatttata agggatgttg
                                                                       1380
ccgatttcgg cctattggtt aaaaaatgag ctgatttaac aaaaatttta acaaaattca
                                                                       1440
                                                                       1500
gaagaacteg teaagaagge gatagaagge gatgegetge gaategggag eggegatace
                                                                       1560
gtaaagcacg aggaagcggt cagcccattc gccgccaagc tettcagcaa tatcacgggt
                                                                       1620
agccaacgct atgtcctgat agcggtccgc cacacccagc cggccacagt cgatgaatcc
agaaaagcgg ccattttcca ccatgatatt cggcaagcag gcatcgccat gggtcacgac
                                                                       1680
                                                                       1740
gagatecteg cegtegggea tgetegeett gageetggeg aacagttegg etggegegag
eccetgatge tettegteca gateatectg ategacaaga eeggetteca teegagtaeg
                                                                       1800
tgctcgctcg atgcgatgtt tcgcttggtg gtcgaatggg caggtagccg gatcaagcgt
                                                                       1860
                                                                       1920
atgcagccgc cgcattgcat cagccatgat ggatactttc tcggcaggag caaggtgaga
tgacaggaga tectgeeceg geacttegee caatageage cagteeette eegetteagt
                                                                       1980
                                                                       2040
gacaacgteg agcacagetg egcaaggaac gecegtegtg gecagecaeg atageegege
 tgcctcgtct tgcagttcat tcagggcacc ggacaggtcg gtcttgacaa aaagaaccgg
                                                                       2100
gegeceetge getgacagee ggaacaegge ggeateagag cageegattg tetgttgtge
                                                                      2160
```

2220

ccagtcatag, ccgaatagcc tctccaccca agcggccgga gaacctgcgt gcaatccatc

```
ttgttcaatc atgcgaaacg atcctcatcc tgtctcttga tcagatcttg atcccctgcg
                                                                      2280
ccatcagatc cttggcggcg agaaagccat ccagtttact ttgcagggct tcccaacctt
                                                                      2340
                                                                      2400
accagagggc gccccagctg gcaattccgg ttcgcttgct gtccataaaa ccgcccagtc
tagctatcgc catgtaagcc cactgcaagc tacctgcttt ctctttgcgc ttgcgttttc
                                                                      2460
                                                                      2520
cettgtecag atageceagt agetgacatt cateeggggt cageacegtt tetgeggact
                                                                      2580
ggctttctac gtgaaaagga tctaggtgaa gatccttttt gataatctca tgaccaaaat
                                                                      2640
cccttaacgt gagttttcgt tccactgagc gtcagacccc gtagaaaaga tcaaaggatc
                                                                      2700
ttettgagat cetttttte tgegegtaat etgetgettg caaacaaaaa aaccaceget
                                                                      2760
accageggtg gtttgtttge eggateaaga getaccaact ettttteega aggtaactgg
                                                                      2820
cttcagcaga gcgcagatac caaatactgt ccttctagtg tagccgtagt taggccacca
cttcaagaac totgtagcac cgcctacata cotogctotg ctaatcotgt taccagtggc
                                                                      2880
tgctgccagt ggcgataagt cgtgtcttac cgggttggac tcaagacgat agttaccgga
                                                                      2940
                                                                      3000
taaggcgcag cggtcgggct gaacgggggg ttcgtgcaca cagcccagct tggagcgaac
                                                                      3060
gacctacacc gaactgagat acctacageg tgagctatga gaaagegeea egetteecga
                                                                      3120
agggagaaag gcggacaggt atccggtaag cggcagggtc ggaacaggag agcgcacgag
ggagcttcca gggggaaacg cctggtatct ttatagtcct gtcgggtttc gccacctctg
                                                                      3180
                                                                      3240
acttgagcgt cgatttttgt gatgctcgtc aggggggggg agcctatgga aaaacgccag
caacgeggee tttttaeggt teetgggett ttgetggeet tttgeteaca tgttetttee
                                                                      3300
tgcgttatcc cctgattctg tggataaccg tattaccgcc tttgagtgag ctgataccgc
                                                                      3360
togoogcago ogaacgacog agogoagoga gtoagtgago gaggaagogg aag
                                                                      3413
```

<210> 11 <211> 4961 <212> DNA

<213> Viral

<400> 11

agegeceaat aegeaaaceg eeteteeeeg egegttggee gatteattaa tgeagetgge 60 acgacaggtt tecegaetgg aaagegggea gtgagegeaa egeaattaat gtgagttage 120 tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa 180 240 ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac gccaagctat ttaggtgaca ctatagaata ctcaagctat gcatcaagct tggtaccgag ctcggatcca 300 360 ctagtaacgg ccgccagtgt gctggaattc atgggcagac ccgtctgtac tttaagagtg 420 ttggcaacca gtaatgaata aaaactcccg ttttattata tttgatgaat gctgaaagct 480 tacattaata tgtcgtgcga tggcacgaaa aaacacacgc aaacaataca ggggggtagt cggcgggcgg ctaagggtgg tgctcggcgg gcagaacatc gaaaaatcaa gatctatatg 540 aattacactt cctccgtagg aggaagcaca gggggagaat accacttctc ccccggcgac 600 660 ataatgtaaa tgacgcagtt tgcctcgaaa tactccagct gccctggagt catttccttc atccaatctt catccgagtt ggcgaggatt attgtaggct tagacttctt ctgcaccttt 720 ttcttcttac catacttggg gtttacaatg aaatccctct gacagccaac taactgtttc 780 caacaaggac agaatttaaa cggaatatca tctacgatgt tgtagattgc gtcttcgttg 840 tatgaagacc aatcaacatt attttgccag taattatgaa cccctaggct tctggcccaa 900 gtagattttc cggttcttgt tgggccgacg atgtagaggc tctgctttct tgatctttca 960 tctgatgact ggatacagaa tccatccatt ggaggtcaga aattgcatcc tcgagggtat 1020 1080 aacaggtagg ttgaaggagc atgtaagctt cgggactaac ctggaagatg ttaggctgga gccaatcgtt gattgactca ttacaaagta aatcaggtga ggagggtgga tgaggattgg 1140 1200 tgaactcttc ctgaatctca ggaaaaagct tatttgcaga gtattcaaaa tactgcaatt 1260 ttgtggacca atcaaagggg agctctttct ggatcatgga gaggtactct tctttggagg tagogtgtga aataatgtot ogcattattt catotttaga aggotttttt tootttacot 1320 ctgaatcaga ttttcctagg aagggggact tcctaggaat gaaagtacct ctctcaaaca 1380 cagccagagg tteettgaga atgtaateee teactetgtt aactgaettg geactetgaa 1440 tatttgggtg aaacccattt atatcaaaga accttgagtc agatatcctt atcggcttct 1500 1560 ctggctgaag caatgcatgt aaatgcaaac ttccatcttt atgtgcctct cgggcacata 1620 gaatatattt gggaatccaa cgaacgacga gctcccagat catctgacag gcgatttcag 1680 gattttctgg acactttgga taggttagga acgtgttagc gttcctgtgt gagaactgac 1740 ggttggatga ggaggaggcc atagccgacg acggaggttg aggctgaggg atggcagact gggageteca aactetatag tataceegtg egeettegaa ateegeeget ecattgtett 1800

atagtggttg	taaatgggcc	ggaccgggcc	ggcccagcag	gaaaagaagg	cgcgcactaa	1860
tattaccgcg	ccttcttttc	ctqcqagggc	ccggtaggga	ccgagcgctt	tgatttaaag	1920
cctggttctg	ctttqtatqa	tttatctaaa	gcagcccaat	ctaaagaaac	cggtcccggg	1980
cactataaat	tocctaacaa	gtgcgattca	ttcatggatc	ctttaaactc	gagtctagag	2040
ggcccaattc	accetataat	gagtcgtatt	acaattcact	ggccgtcgtt	ttacaacgtc	2100
gtgactggga	aaaccctggc	gttacccaac	ttaatcgcct	tgcagcacat	cccctttcg	2160
ccaactaaca	taatagcgaa	gaggcccgca	ccgatcgccc	ttcccaacag	ttgcgcagcc	2220
tatacqtacq	gcagtttaag	gtttacacct	ataaaagaga	gagccgttat	cgtctgtttg	2280
togatgtaca	gagtgatatt	attgacacge	cggggcgacg	gatggtgatc	cccctggcca	2340
gracacatet	octotcagat	aaagtctccc	gtgaacttta	cccggtggtg	catatcgggg	2400
atgaaagctg	gcgcatgatg	accaccgata	tggccagtgt	gccggtctcc	gttatcgggg	2460
aagaagtggc	toatctcage	caccgcgaaa	atgacatcaa	aaacgccatt	aacctgatgt	2520
totagggaat	ataaatgtca	gacctgaatg	gcgaatggac	gcgccctgta	gcggcgcatt	2580
aaacacacaa	atataataat	tacqcqcaqc	gtgaccgcta	cacttgccag	cgccctagcg	2640
cccactcctt	tegetttett	cccttccttt	ctcgccacgt	tegeeggett	tccccgtcaa	2700
gctctaaatc	aaaaactccc	tttagggttc	cgatttagag	ctttacggca	cctcgaccgc	2760
aaaaaacttq	atttgggtga	tagttcacqt	agtgggccat	cgccctgata	gacggttttt	2820
caccettaa	cattagaatc	cacqttcttt	aatagtggac	tettgtteca	aactggaaca	2880
acactcaacc	ctatcgcggt	ctattctttt	gatttataag	ggatgttgcc	gatttcggcc	2940
tattoottaa	aaaatgagct	gatttaacaa	aaattttaac	aaaattcaga	agaactcgtc	3000
aagaaggcga	tagaaggcga	tacactacaa	atcgggagcg	gcgataccgt	aaagcacgag	3060
gaagcggtca	gcccattcgc	caccaaacte	ttcagcaata	tcacgggtag	ccaacgctat	3120
gtcctgatag	caatccacca	cacccagccg	gccacagtcg	atgaatccag	aaaagcggcc	3180
attttccacc	atgatattcg	gcaagcaggc	atcgccatgg	gtcacgacga	gatcctcgcc	3240
atcagacata	ctcqccttqa	gcctggcgaa	cagttcggct	ggcgcgagcc	cctgatgctc	3300
ttcatccaga	tcatcctgat	cgacaagacc	ggcttccatc	cgagtacgtg	ctcgctcgat .	3360
acaatatttc	acttaataat	cgaatgggca	ggtagccgga	tcaagcgtat	gcagccgccg	3420
cattocatca	gccatgatgg	atactttctc	ggcaggagca	aggtgagatg	acaggagatc	3480
ctaccccaac	acttcqccca	atagcagcca	gtcccttccc	gcttcagtga	caacgtcgag	3540
cacagetgeg	caaggaacgc	ccgtcgtggc	cagccacgat	agecgegetg	cctcgtcttg	3600
caqttcattc	agggcaccgg	acaggtcggt	cttgacaaaa	agaaccgggc	gcccctgcgc	3660
tgacagccgg	aacacggcgg	catcagagca	gccgattgtc	tgttgtgccc	agtcatagcc	3720
qaataqcctc	tccacccaag	cggccggaga	acctgcgtgc	aatccatctt	gttcaatcat	3780
gcgaaacgat.	cctcatcctg	tctcttgatc	agatcttgat	cccctgcgcc	atcagatcct.	3840
taacaacaaa	aaaqccatcc	agtttacttt	gcagggcttc	ccaaccttac	cagagggcgc	′3900
cccaqctqqc	aattccggtt	cgcttgctgt	ccataaaacc	gcccagtcta	gctatcgcca	3960
tataaaccca	ctqcaaqcta	cctqctttct	ctttgcgctt	gcgttttccc	ttgtccagat	4020
agcccagtag	ctgacattca	tccggggtca	gcaccgtttc	tgcggactgg	ctttctacgt	4080
qaaaaqqatc	taggtgaaga	teettttiga	taatctcatg	accaaaatcc	cttaacgtga	4140
gttttcgttc	cactgagcgt	cagaccccgt	agaaaagatc	aaaggatctt	cttgagatcc	4200
tttttttctg	cgcgtaatct	gctgcttgca	aacaaaaaa	ccaccgctac	cagcggtggt	4260
ttatttacca	gatcaagagc	taccaactct	ttttccgaag	gtaactggct	tcagcagagc	4320
gcagatacca	aatactgtcc	ttctagtgta	gccgtagtta	ggccaccact	tcaagaactc	4380
tgtagcaccg	cctacatacc	tegetetget	aatcctgtta	ccagtggctg	ctgccagtgg	4440
cgataagtcg	tgtcttaccg	ggttggactc	aagacgatag	ttaccggata	aggcgcagcg	4500
gtcgggctga	acggggggtt	cgtgcacaca	gcccagcttg	gagcgaacga	cctacaccga	4560
actgagatac	ctacagcgtg	agctatgaga	aagcgccacg	cttcccgaag	ggagaaaggc	4620
ggacaggtat	ccggtaagcg	gcagggtcgg	aacaggagag	cgcacgaggg	agcttccagg	4680
gggaaacgcc	tggtatcttt	atagtcctgt	cgggtttcgc	cacctctgac	ttgagcgtcg	4740
atttttqtqa	tgctcgtcag	gggggcggag	cctatggaaa	aacgccagca	acgcggcctt	4800
tttacggttc	ctgggctttt	gctggccttt	tgctcacatg	ttctttcctg	cgttatcccc	4860
tgattctgtg	gataaccgta	ttaccgcctt	tgagtgagct	gataccgctc	gccgcagccg	4920
			ggaagcggaa			4961
	- *					

<210> 12

<211> 6309

<212> DNA

<213> Viral

<400> 12 60 agegeceaat aegeaaaeeg ceteteeeeg egegttggee gatteattaa tgeagetgge acgacaggtt tcccgactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc 120 tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa. 180 ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac gccaagctat 240 ttaggtgaca ctatagaata ctcaagctat gcatcaagct tggtaccgag ctcggatcca 300 360 ctagtcccga tctagtaaca tagatgacac cgcgcgcgat aatttatcct agtttgcgcg 420 ctatattttg ttttctatcg cgtattaaat gtataattgc gggactctaa tcataaaaac ccatctcata aataacgtca tgcattacat gttaattatt acatgcttaa cgtaattcaa 480 540 cagaaattat atgataatca tcgacagacc ggcaacagga ttcaatctta agaaacttta ttgccaaatg tttgaacgat cggggaaatt cgctcgagtt aattaagcgg ccgcctcaaa 600 660 aaggatette acctagatee ttttaaatta aaaatgaagt tttagcaegt gteagteetg 720 ctcctcggcc acgaagtgca cgcagttgcc ggccgggtcg cgcagggcga actcccgccc 780 ccacggetge tegeogatet eggteatgge eggeoeggag gegteeegga agttegtgga cacgacetee gaccactegg egtacagete gtecaggeeg egeacecaca eccaggeeag 840 900 ggtgttgtcc ggcaccacct ggtcctggac cgcgctgatg aacagggtca cgtcgtcccg gaccacaccg gcgaagtcgt cctccacgaa gtcccgggag aacccgagcc ggtcggtcca 960 gaactcgace gctccggcga cgtcgcgcgc ggtgagcacc ggaacggcac tggtcaactt 1020 ggccatggtg gccctcctca cgtgctatta ttgaagcatt tatcagggtt attgtctcat 1080 gageggatac atatttgaat gtatttagaa aaataaacaa ataggggtte egegeacatt 1140 1200 tccccgaaaa gtgccacctg tatgcggtgt gaaataccgc acagatgcgt aaggagaaaa 1260 taccgcatca ggcgaaattg taaacgcggc cgcttaatta agtcgacgtc ctctccaaat gaaatgaact teettatata gaggaagggt ettgegaagg atagtgggat tgtgegteat 1320 1380 cccttacgtc agtggagata tcacatcaat ccacttgctt tgaagacgtg gttggaacgt cttctttttc cacgtagctc ctcgtgggtg ggggtccatc tttgggacca ctgtcggcag 1440 aggcatcttg aacgatagcc tttccttatc gcaatgatgg catttgtagg tgccaccttc 1500 cttttctact gtccttttga tgaagtgaca gatagctggg caatggaatc cgaggaggtt 1560 1620 tecegatatt accettigtt gaaaagtete aatageeett tggtettetg agaetgtate 1680 tttgatattc ttggagtaga cgagagagtg tcgtgctcca ccatgttgac gaattcatgg gcagacccgt ctgtacttta agagtgttgg caaccagtaa tgaataaaaa ctcccgtttt 1740 1800 attatatttg atgaatgctg aaagcttaca ttaatatgtc gtgcgatggc acgaaaaaac 1860 acacgcaaac aatacagggg ggtagtcggc gggcggctaa gggtggtgct cggcgggcag 1920 aacatcgaaa aatcaagatc tatatgaatt acacttcctc cgtaggagga agcacagggg 1980 gagaatacca cttctccccc ggcgacataa tgtaaatgac gcagtttgcc tcgaaatact ccagctgccc tggagtcatt tccttcatcc aatcttcatc cgagttggcg aggattattg 2040 taggettaga ettettetge acetttttet tettaceata ettggggttt acaatgaaat 2100 2160 ccctctgaca gccaactaac tgtttccaac aaggacagaa tttaaacgga atatcatcta cgatgttgta gattgcgtct tcgttgtatg aagaccaatc aacattattt tgccagtaat 2220 2280 tatgaacccc taggettetg geceaagtag atttteeggt tettgttggg eegaegatgt . 2340 agaggetetg etttettgat ettteatetg atgaetggat acagaateea tecattggag gtcagaaatt gcatcetega gggtataaca ggtaggttga aggagcatgt aagetteggg 2400 2460 actaacctgg aagatgttag gctggagcca atcgttgatt gactcattac aaagtaaatc 2520 aggtgaggag ggtggatgag gattggtgaa ctcttcctga atctcaggaa aaagcttatt tgcagagtat tcaaaatact gcaattttgt ggaccaatca aaggggagct ctttctggat 2580 2640 catggagagg tactcttctt tggaggtagc gtgtgaaata atgtctcgca ttatttcatc tttagaaggc tttttttcct ttacctctga atcagatttt cctaggaagg gggacttcct 2700 aggaatgaaa gtacctctct caaacacagc cagaggttcc ttgagaatgt aatccctcac 2760 2820 tctgttaact gacttggcac tctgaatatt tgggtgaaac ccatttatat caaagaacct tgagtcagat atccttatcg gcttctctgg ctgaagcaat gcatgtaaat gcaaacttcc 2880 atctttatgt gcctctcggg cacatagaat atatttggga atccaacgaa cgacgagctc 2940 ccagatcatc tgacaggcga tttcaggatt ttctggacac tttggatagg ttaggaacgt 3000 gttagegtte etgtgtgaga aetgaeggtt ggatgaggag gaggeeatag eegaegaegg 3060 aggttgaggc tgagggatgg cagactggga gctccaaact ctatagtata cccgtgcgcc 3120 ttcgaaatcc gccgctccat tgtcttatag tggttgtaaa tgggccggac cgggccggcc 3180 3240 cagcaggaaa agaaggcgcg cactaatatt accgcgcctt cttttcctgc gagggcccgg

						2200
ggtagggacc	gagcgctttg	atttaaagcc	tggttctgct	ttgtatgatt	tatctaaagc	3300
agcccaatct	aaagaaaccg	gtcccgggca	ctataaattg	cctaacaagt	gcgattcatt	3360
catggatect	ttaaactcga	gtctagaggg	cccaattcgc	cctatagtga	gtcgtattac	3420
aattcactgg	ccgtcgtttt	acaacgtcgt	gactgggaaa	accetggegt	tacccaactt	3480
aatcgccttg	cagcacatcc	ccctttcgcc	agctggcgta	atagcgaaga	ggcccgcacc	3540
gatcgccctt	cccaacagtt	gcgcagccta	tacgtacggc	agtttaaggt	ttacacctat	3600
aaaagagaga	gccgttatcg	tctgtttgtg	gatgtacaga	gtgatattat	tgacacgccg	3660
gggcgacgga	tggtgatccc	cctggccagt	gcacgtctgc	tgtcagataa	agteteeegt	3720
qaactttacc	cggtggtgca	tatcggggat	gaaagctggc	gcatgatgac	caccgatatg	3780
gccagtgtgc	cggtctccgt	tatcggggaa	gaagtggctg	atctcagcca	ccgcgaaaat	3840
gacatcaaaa	acgccattaa	cctgatgttc	tggggaatat	aaatgtcagg	cctgaatggc	3900
gaatggacgc	gccctgtagc	ggcgcattaa	gcgcgcgggt	gtggtggtta	cgcgcagcgt	3960
gaccgctaca	cttgccagcg	ccctagcgcc	cgctcctttc	gctttcttcc	CTTCCTTCC	4020
caccacattc	gccggctttc	cccgtcaagc	tctaaatcgg	gggctccctt	tagggttccg	4080
atttagagct	ttacggcacc	tcgaccgcaa	aaaacttgat	ttgggtgatg	gttcacgtag	4140
tgggccatcg	ccctgataga	cggtttttcg	ccctttgacg	ttggagtcca	cgttctttaa	4200
tagtggactc	ttqttccaaa	ctggaacaác	actcaaccct	atcgcggtct	attcttttga	4260
tttataaqqq	atgttgccga	tttcggccta	ttggttaaaa	aatgagctga	tttaacaaaa	4320
attttaacaa	aattcagaag	aactcgtcaa	gaaggcgata	gaaggcgatg	cgctgcgaat	4380
caggaggggc	gataccgtaa	agcacgagga	agcggtcagc	ccattcgccg	ccaagctctt	4440
cagcaatatc	acqqqtagcc	aacgctatgt	cctgatagcg	gtccgccaca	cccagccggc	4500
cacagtcgat	qaatccaqaa	aagcggccat	tttccaccat	gatattcggc	aagcaggcat	4560
caccatagat	cacqacqaqa	tectegeegt	cgggcatgct	cgccttgagc	ctggcgaaca	4620
atteaactaa	cacaaacccc	tgatgctctt	cgtccagatc	atcctgatcg	acaagaccgg	4680
cffccatccg	agtacgtgct	cgctcgatgc	gatgtttcgc	ttggtggtcg	aatgggcagg	4740
-tagccggatc	aagcgtatgc	agccgccgca	ttgcatcagc	catgatggat	actttctcgg	4800
caggaggaag	gtgagatgac	aggagatect	qccccggcac	ttcgcccaat	agcagccagt	4860
cccttcccac	ttcagtgaca	acgtcgagca	cagctgcgca	aggaacgccc	gtcgtggcca	4920
accacgatag	ccacactacc	tegtettgca	gttcattcag	ggcaccggac	aggtcggtct	4980
tgacaaaaag	aaccaaacac	ccctgcgctg	acageeggaa	cacggcggca	tcagagcagc	5040
casttateta	ttgtgcccag	tcatagccga	atageetete	cacccaagcg	gccggagaac	5100
ctacatacaa	tecatettet	tcaatcatgc	gaaacgatcc	tcatcctgtc	tcttgatcag	5160
atcttgatcc	cctgcgccat	cagatccttg	gcqqcqaqaa	agccatccag	tttactttgc	5220
agggetteec	aaccttacca	gagggcgccc	cagetggeaa	ttccqgttcg	cttgctgtcc	5280
araaaaccac	ccagtctagc	tatcgccatg	taagcccact	gcaagctacc	tgctttctct	5340
tracacttac	attttccctt	gtccagatag	cccaqtaqct	qacattcatc	cggggtcagc	5400
accontition	cagactaget	ttctacgtga	aaaggatcta	ggtgaagatc	ctttttgata	5460
atctcatoac	caaaatccct	taacgtgagt	tttcqttcca	ctgagcgtca	gaccccgtag	5520
aaaagatgaa	aggatettet	tgagatcctt	tttttctaca	cgtaatctgc	tgcttgcaaa	5580
caaaaaaacc	accoctacca	gcggtggttt	atttaccaga	tcaagagcta	ccaactcttt	5640
ttcccaacat	aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	ctagtgtagc	5700
cataattaaa	ccaccacttc	aagaactctg	tagcaccgcc	tacatacctc	gctctgctaa	57 6 0
tectettace	agtogctoct	gccagtggcg	ataagtcgtg	tcttaccggg	ttggactcaa	5820
gacgatagt	acconstance	gcgcagcggt	coogctgaac	agagaattca	tgcacacagc	5880
gacgatagee	accesaccec	tacaccgaac	tgagatacct	acagogtgag	ctatgagaaa	5940
acaccacact	tcccasagg	adaaaaacaa	acaggtatee	ggtaagcggc	agggtcggaa	6000
gegeeaeger	caccagaayyy	cttccadaga	gaaacgcctg	gtatctttat	agtcctgtcg	6060
cayyayaycy	catchagggag	caccagggg	ttttataata	ctcatcagga	gggcggagcc	6120
ggtttegeca	caccacacaca	gagegeegat	tacoettect	gaacttttac	tggccttttg	6180
catggaaaaa	otttostss	ttatcccctc	attetatace	taaccotatt	accgcctttg	6240
ctcacatgtt	taggestess	cacacacaca	cascasaca	cadcaagtca	gtgagcgagg	6300
	Lacegerege	cycayccyaa	carccaraca	3-3-500		6309
aagcggaag						3.0.0

<210> 13 <211> 8043

<212> DNA <213> Viral

<400> 13 agogoccaat acgcaaacog cototocoog ogogttggoo gattoattaa tgcagotggo 60 120 acgacaggtt tecegactgg aaagegggea gtgagegeaa egeaattaat gtgagttage 180 tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa 240 ttgtgagegg ataacaattt cacacaggaa acagetatga ecatgattac gecaagetat 300 ttaggtgaca ctatagaata ctcaagctat gcatcaagct tggtaccgag ctcggatcca ctagtaacgg ccgccagtgt gctggaattc atgggcagac ccgtctgtac tttaagagtg 360 420 ttggcaacca gtaatgaata aaaactcccg ttttattata tttgatgaat gctgaaagct tacattaata tgtcgtgcga tggcacgaaa aaacacacgc aaacaataca ggggggtagt 480 540 cggcgggcgg ctaagggtgg tgctcggcgg gcagaacatc gaaaaatcaa gatctatatg 600 aattacactt cctccgtagg aggaagcaca gggggagaat accacttctc ccccggcgac 660 ataatgtaaa tgacgcagtt tgcctcgaaa tactccagct gccctggagt catttccttc atccaatctt catccgagtt ggcgaggatt attgtaggct tagacttctt ctgcaccttt 720 780 ttottottac catacttggg gtttacaatg aaatccotot gacagccaac taactgttto caacaaggac agaatttaaa cggaatatca tctacgatgt tgtagattgc gtcttcgttg 840 tatgaagacc aatcaacatt attttgccag taattatgaa cccctaggct tctggcccaa 900 gtagattttc cggttcttgt tgggccgacg atgtagaggc tctgctttct tgatctttca 960 1020 totgatgact ggatacagaa tocatocatt ggaggtoaga aattgcatoo togagggtat aacaggtagg ttgaaggagc atgtaagctt cgggactaac ctggaagatg ttaggctgga 1080 gccaatcgtt gattgactca ttacaaagta aatcaggtga ggagggtgga tgaggattgg 1140 tgaactette etgaatetea ggaaaaaget tatttgeaga gtatteaaaa taetgeaatt 1200 ttgtggacca atcaaagggg agctctttct ggatcatgga gaggtactct tctttggagg 1260 tagegtgtga aataatgtet egeattattt catetttaga aggetttttt teetttaeet 1320 ctgaatcaga ttttcctagg aagggggact tcctaggaat gaaagtacct ctctcaaaca 1380 cagccagagg ttccttgaga atgtaatccc tcactctgtt aactgacttg gcactctgaa 1440 tatttgggtg aaacccattt atatcaaaga accttgagtc agatatcctt atcggcttct 1500 ctggctgaag caatgcatgt aaatgcaaac ttccatcttt atgtgcctct cgggcacata 1560 1620 gaatatattt gggaatccaa cgaacgacga gctcccagat catctgacag gcgatttcag 1680 gattttctgg acactttgga taggttagga acgtgttagc gttcctgtgt gagaactgac 1740 ggttggatga ggaggaggcc atagccgacg acggaggttg aggctgaggg atggcagact 1800 gggageteca aactetatag tataccegtg egeettegaa ateegeeget ceattgtett atagtggttg taaatgggcc ggaccgggcc ggcccagcag gaaaagaagg cgcgcactaa 1860 tattaccgcg ccttcttttc ctgcgagggc ccggtaggga ccgagcgctt tgatttaaag 1920 1980 cctggttctg ctttgtatga tttatctaaa gcagcccaat ctaaagaaac cggtcccggg 2040 cactataaat tgcctaacaa gtgcgattca ttcatggatc ctttaaactc gagtctagtc ccgatctagt aacatagatg acaccgcgcg cgataattta tcctagtttg cgcgctatat 2100 tttgttttct atcgcgtatt aaatgtataa ttgcgggact ctaatcataa aaacccatct 2160 2220 cataaataac gtcatgcatt acatgttaat tattacatgc ttaacgtaat tcaacagaaa 2280 ttatatgata atcatcgaca gaccggcaac aggattcaat cttaagaaac tttattgcca 2340 aatgtttgaa cgatcgggga aattcgctcg agttaattaa gcggccgcct caaaaaggat cttcacctag atccttttaa attaaaaatg aagttttagc acgtgtcagt cctgctcctc 2400 ggccacgaag tgcacgcagt tgccggccgg gtcgcgcagg gcgaactccc gcccccacgg 2460 ctgctcgccg atctcggtca tggccggccc ggaggcgtcc cggaagttcg tggacacgac 2520 ctccgaccac tcggcgtaca gctcgtccag gccgcgcacc cacacccagg ccagggtgtt 2580 2640 gtccggcacc, acctggtcct ggaccgcgct gatgaacagg gtcacgtcgt cccggaccac 2700 accggcgaag tcgtcctcca cgaagtcccg ggagaacccg agccggtcgg tccagaactc gaccgctccg gcgacgtcgc gcgcggtgag caccggaacg gcactggtca acttggccat 2760 2820 ggtggccctc ctcacgtgct attattgaag catttatcag ggttattgtc tcatgagcgg 2880 atacatattt gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg 2940 aaaagtgcca cctgtatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc 3000 3060 aacttcctta tatagaggaa gggtcttgcg aaggatagtg ggattgtgcg tcatccctta 3120 cgtcagtgga gatatcacat caatccactt gctttgaaga cgtggttgga acgtcttctt tttccacgta gctcctcgtg ggtgggggtc catctttggg accactgtcg gcagaggcat 3180

```
cttgaacgat agcctttcct tatcgcaatg atggcatttg taggtgccac cttccttttc
                                                                      3240
tactgtcctt ttgatgaagt gacagatagc tgggcaatgg aatccgagga ggtttcccga
                                                                      3300
tattaccett tgttgaaaag teteaatage eetttggtet tetgagaetg tatetttgat
                                                                      3360
                                                                      3420
attottggag tagacgagag agtgtcgtgc tocaccatgt tgacgaattc atgggcagac
                                                                      3480
cogtotgtac tttaagagtg ttggcaacca gtaatgaata aaaactcccg ttttattata
                                                                      3540
tttgatgaat gctgaaagct tacattaata tgtcgtgcga tggcacgaaa aaacacacgc
                                                                      3600
aaacaataca ggggggtagt cggcgggcgg ctaagggtgg tgctcggcgg gcagaacatc
                                                                      3660
gaaaaatcaa gatctatatg aattacactt cctccgtagg aggaagcaca gggggagaat
accacttctc ccccggcgac ataatgtaaa tgacgcagtt tgcctcgaaa tactccagct
                                                                      3720
                                                                      3780
geoetggagt cattteette atecaatett cateegagtt ggegaggatt attgtagget
tagacttett etgeacettt ttettettae catacttggg gtttacaatg aaateeetet
                                                                      3840
gacagecaae taaetgttte caacaaggae agaatttaaa eggaatatea tetaegatgt
                                                                      3900
tgtagattgc gtcttcgttg tatgaagacc aatcaacatt attttgccag taattatgaa
                                                                      3960
cccctaggct tctggcccaa gtagattttc cggttcttgt tgggccgacg atgtagaggc
                                                                      4020
                                                                      4080
tetgetttet tgatetttea tetgatgaet ggatacagaa tecatecatt ggaggteaga
aattgcatcc tcgagggtat aacaggtagg ttgaaggagc atgtaagctt cgggactaac
                                                                      4140
                                                                      4200
ctggaagatg ttaggctgga gccaatcgtt gattgactca ttacaaagta aatcaggtga
ggagggtgga tgaggattgg tgaactette etgaatetea ggaaaaaget tatttgeaga
                                                                      4260
gtattcaaaa tactgcaatt ttgtggacca atcaaagggg agctctttct ggatcatgga
                                                                      4320
gaggtactct tctttggagg tagcgtgtga aataatgtct cgcattattt catctttaga
                                                                      4380
aggetttttt teetttaeet etgaateaga tttteetagg aagggggaet teetaggaat
                                                                      4440
                                                                      4500
gaaagtacct ctctcaaaca cagccagagg ttccttgaga atgtaatccc tcactctgtt
aactgacttg gcactctgaa tatttgggtg aaacccattt atatcaaaga accttgagtc
                                                                      4560
agatateett ateggettet etggetgaag caatgeatgt aaatgeaaae tteeatettt
                                                                      4620
                                                                      4680
atgtgcctct cgggcacata gaatatattt gggaatccaa cgaacgacga gctcccagat
catctgacag gcgatttcag gattttctgg acactttgga taggttagga acgtgttagc
                                                                      4740
gttcctgtgt gagaactgac ggttggatga ggaggaggcc atagccgacg acggaggttg
                                                                      4800
aggetgaggg atggeagact gggageteca aactetatag tataccegtg egeettegaa
                                                                      4860
                                                                      4920
atccgccgct ccattgtctt atagtggttg taaatgggcc ggaccgggcc ggcccagcag
                                                                      4980
gaaaagaagg cgcgcactaa tattaccgcg ccttcttttc ctgcgagggc ccggggtagg
gaccgagcgc tttgatttaa agcctggttc tgctttgtat gatttatcta aagcagccca
                                                                      5040
atctaaagaa accggtcccg ggcactataa attgcctaac aagtgcgatt cattcatgga
                                                                      5100
                                                                      5160
teetttaaae tegagtetag agggeecaat tegeeetata gtgagtegta ttacaattea
ctggccgtcg ttttacaacg tcgtgactgg gaaaaccctg gcgttaccca acttaatcgc
                                                                      5220
                                                                      5280
cttgcagcac atcccccttt cgccagctgg cgtaatagcg aagaggcccg caccgatcgc
ccttcccaac agttgcgcag cctatacgta cggcagttta aggtttacac ctataaaaga
                                                                      5340
                                                                      5400
gagageegtt ategtetgtt tgtggatgta cagagtgata ttattgacae geeggggega
                                                                      5460
cggatggtga tececetgge cagtgeacgt etgetgteag ataaagtete eegtgaactt
                                                                      5520
tacccggtgg tgcatatcgg ggatgaaagc tggcgcatga tgaccaccga tatggccagt
                                                                       5580
gtgccggtct ccgttatcgg ggaagaagtg gctgatctca gccaccgcga aaatgacatc
aaaaacgcca ttaacctgat gttctgggga atataaatgt caggcctgaa tggcgaatgg
                                                                       5640
acgcgccctg tagcggcgca ttaagcgcgc gggtgtggtg gttacgcgca gcgtgaccgc
                                                                       5700
tacacttgcc agogecetag egecegetee tttegettte tteeetteet ttetegecac
                                                                       5760
gttcgccggc tttccccgtc aagctctaaa tcgggggctc cctttagggt tccgatttag
                                                                       5820
agetttaegg cacetegace gcaaaaaact tgatttgggt gatggttcae gtagtgggee
                                                                       5880
ategecetga tagaeggttt ttegecettt gaegttggag teeaegttet ttaatagtgg
                                                                       5940
                                                                       6000
actettgtte caaactggaa caacacteaa ceetategeg gtetattett ttgatttata
agggatgttg ccgatttcgg cctattggtt aaaaaatgag ctgatttaac aaaaatttta
                                                                       6060
                                                                       6120
acaaaattca gaagaactcg tcaagaaggc gatagaaggc gatgcgctgc gaatcgggag
cggcgatacc gtaaagcacg aggaagcggt cagcccattc gccgccaagc tcttcagcaa
                                                                       6180
 tatcacgggt agccaacgct atgtcctgat agcggtccgc cacacccagc cggccacagt
                                                                       6240
cgatgaatcc agaaaagcgg ccattttcca ccatgatatt cggcaagcag gcatcgccat
                                                                       6300
 gggtcacgac gagatecteg cegtegggea tgetegeett gageetggeg aacagttegg
                                                                       6360
                                                                       6420
 ctggcgcgag cccctgatgc tcttcgtcca gatcatcctg atcgacaaga ccggcttcca
 tecgagtacg tgetegeteg atgegatgtt tegettggtg gtegaatggg caggtageeg .
                                                                       6480
 gatcaagegt atgeageege egeattgeat cagecatgat ggataettte teggeaggag
                                                                       6540
                                                                       6600
 caaggtgaga tgacaggaga teetgeeeeg geaettegee caatageage cagteeette
```

```
ccgcttcagt gacaacgtcg agcacagctg cgcaaggaac gcccgtcgtg gccagccacg
                                                                     6660
atageogogo tgeotogtot tgeagtteat teagggeace ggacaggteg gtettgacaa
                                                                     6720
aaagaaccgg gcgcccctgc gctgacagcc ggaacacggc ggcatcagag cagccgattg
                                                                     6780
totgttgtgc ccagtcatag ccgaatagcc totccaccca ageggeegga gaacetgegt
                                                                     6840
                                                                     6900
gcaatccatc ttgttcaatc atgcgaaacg atcctcatcc tgtctcttga tcagatcttg
atcccctgcg ccatcagatc cttggcggcg agaaagccat ccagtttact ttgcagggct
                                                                     6960
teccaacett accagagge gececagetg geaatteegg ttegettget gtecataaaa
                                                                     7020
                                                                     7080
ccgcccagtc tagctatcgc catgtaagcc cactgcaagc tacctgcttt ctctttgcgc
                                                                     7140
ttgcgttttc ccttgtccag atagcccagt agctgacatt catccggggt cagcaccgtt
tctgcggact ggctttctac gtgaaaagga tctaggtgaa gatccttttt gataatctca
                                                                      7200
tgaccaaaat cccttaacgt gagttttcgt tccactgagc gtcagacccc gtagaaaaga
                                                                      7260
tcaaaggatc ttcttgagat ccttttttc tgcgcgtaat ctgctgcttg caaacaaaaa
                                                                      7320
                                                                     7380
aaccaccgct accagcggtg gtttgtttgc cggatcaaga gctaccaact ctttttccga
aggtaactgg cttcagcaga gcgcagatac caaatactgt ccttctagtg tagccgtagt
                                                                     7440
taggccacca cttcaagaac totgtagcac cgcctacata cotcgctctg ctaatcctgt
                                                                      7500
taccagtggc tgctgccagt ggcgataagt cgtgtcttac cgggttggac tcaagacgat
                                                                      7560
agttaccgga taaggcgcag cggtcgggct gaacgggggg ttcgtgcaca cagcccagct
                                                                      7620
tggagcgaac gacctacacc gaactgagat acctacagcg tgagctatga gaaagcgcca
                                                                      7680
                                                                      7740
cgcttcccga agggagaaag gcggacaggt atccggtaag cggcagggtc ggaacaggag
agcgcacgag ggagcttcca gggggaaacg cctggtatct ttatagtcct gtcgggtttc
                                                                      7800
                                                                      7860
gccacctctg acttgagcgt cgatttttgt gatgctcgtc aggggggggg agcctatgga
aaaacgccag caacgcggcc tttttacggt tcctgggctt ttgctggcct tttgctcaca
                                                                      7920
                                                                      7980
tgttctttcc tgcgttatcc cctgattctg tggataaccg tattaccgcc tttgagtgag
                                                                      8040
ctgatacege tegeogeage egaacgaeeg agegeagega gteagtgage gaggaagegg
                                                                      8043
```

<210> 14 <211> 7404 <212> DNA <213> Viral

<400> 14

agogoccaat acgcaaaccg cototocccg cgcgttggcc gattcattaa tgcagctggc acgacaggtt tcccgactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc 120 tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa 180 ttgtgagegg ataacaattt cacacaggaa acagctatga ccatgattac gccaagctat 240 300 ttaggtgaca ctatagaata ctcaagctat gcatcaagct tggtaccgag ctcggatcca ctagtaacgg ccgccagtgt gctggaattc atgggcagac ccgtctgtac tttaagagtg 360 420 ttggcaacca gtaatgaata aaaactcccg ttttattata tttgatgaat gctgaaagct 480 tacattaata tgtcgtgcga tggcacgaaa aaacacacgc aaacaataca ggggggtagt cggcgggcgg ctaagggtgg tgctcggcgg gcagaacatc gaaaaatcaa gatctatatg 540 aattacactt cctccgtagg aggaagcaca gggggagaat accacttctc ccccggcgac 600 ataatgtaaa tgacgcagtt tgcctcgaaa tactccagct gccctggagt catttccttc 660 720 atccaatctt catccgagtt ggcgaggatt attgtaggct tagacttctt ctgcaccttt 780 ttcttcttac catacttggg gtttacaatg aaatccctct gacagccaac taactgtttc caacaaggac agaatttaaa cggaatatca tctacgatgt tgtagattgc gtcttcgttg 840 tatgaagacc aatcaacatt attttgccag taattatgaa cccctaggct tctggcccaa 900 960 gtagattttc cggttcttgt tgggccgacg atgtagaggc tctgctttct tgatctttca 1020 totgatgact ggatacagaa tocatocatt ggaggtcaga aattgcatco togagggtat aacaggtagg ttgaaggagc atgtaagctt cgggactaac ctggaagatg ttaggctgga 1080 1140 gccaatcgtt gattgactca ttacaaagta aatcaggtga ggagggtgga tgaggattgg 1200 tgaactette etgaatetea ggaaaaaget tatttgeaga gtatteaaaa taetgeaatt ttgtggacca atcaaagggg agctctttct ggatcatgga gaggtactct tctttggagg 1260 tagogtgtga aataatgtot ogcattattt catotttaga aggotttttt tootttacot 1320 ctgaatcaga ttttcctagg aagggggact tcctaggaat gaaagtacct ctctcaaaca 1380

	ttccttgaga	atotaatoco	tractitatt	aactgacttg	gcactctgaa	1440
cagccagagg	assecratt	atatcaaaga	accttgagtc	agatateett	atcggcttct	1500
cattegggtg	caatgcatgt	acatocaaac	ttccatcttt	atgtgcctct	caacata	1560
ctggctgaag	gggaatccaa	casscascas	actoccagat	catctgacag	gcgatttcag	1620
gaatatattt	acactttgga	taggttagga	acgtottage	attectatat	gagaactgac	1680
gatttttttgg	ggaggaggcc	ataggccagga	acquagatta	aggetgaggg	atggcagact	1740
ggttggatga	aactctatag	tatacccata	cacattagaa	atccccccct	ccattgtctt	1800
gggageteea	taaatgggcc	catacceges	adcccaacaa	gaaaagaagg	cacacactaa	1860
acagigging	ccttcttttc	ggaccgggcc ctcccaccc	ccaataggga	ccgagcgctt	tgatttaaag	1920
cattactgcg	ctttgtatga	tttatctaaa	gcagccaat	ctaaagaaac	caatcccaaa	1980
ectggttetg	tgcctaacaa	atacaettce	treategate	ctttaaactc	gagtctagtc	2040
Cactataaat	aacatagatg	acaccacaca	ccataattta	tectagette	cococtatat	2100
ccgatctagt	atcgcgtatt	acaccycycy,	ttacagaact	ctaatcataa	aaacccatct	2160
tttgtttttt	gtcatgcatt	adatgtataat	tattacator	ttaacgtaat	tcaacagaaa	2220
cataaataac	atcatcgaca	acatguraac	accattcaat	cttaagaaac	tttattgcca	2280
ttatatgata	accategaca	gattggtaat	aggacccaac	acaaccactt	aattaagtcg	2340
aatgtttgaa	cgatcgggga	aaccegeeeg	atatagagga	agggtegtet	gaaggatagt	2400
acgreerere	caaatgaaat	gaacttectt	acatatcaca	tcaatccact	toctttgaag	2460
gggattgtgc	gtcatccctt	ttttaaaa	agatactaca	agatagaat	ccatctttqq	2520
acgtggttgg	aacgtcttct	tettecacge,	tagecettee	ttatcgcaat	gatggcattt	2580
gaccactgtc	ggcagaggca	stactatest	tttgatgag	tracaratar	ctaggcaata	2640
graggracca	ccttcctttt	atattaggg	ttattasss	gtctcaatag	ccctttaatc	2700
gaateegagg	aggtttcccg	tottattage	at acaccaca	gactatag	ctccaccatg	
ttctgagact	gtatetttga	caccettgga	grayargaga	gagegeegeg	ctccaccatg	2820
ttgacgaatt	catgggcaga	ceegtetgta	tactassage	ttacattaat	atatcataca	2880
aaaaactccc	gttttattat	accigacyaa	rgctgaaagc	teggegggg	actaaggata	2940
atggcacgaa	aaaacacacg	caaacaatac	aggggggtag	ceggegggeg	tectecataa	3000
gtgctcggcg	ggcagaacat	cgaaaaatca	agatetatat	cataatotaa	atgacgcagt	3060
gaggaagcac	agggggagaa	taccactict	teattteett	cataatgtaa	tcatccgagt	3120
ttgcctcgaa	atactccagc	tgecetggag	totacecet	tttcttctta	ccatacttgg	3180
tggcgaggat	tattgtaggc	tragactict	ctaactett	ccaacaagga	cagaatttaa	3240
ggtttacaat	gaaatccctc	tgacagecaa	ccaactgett	gtatgaagac	caatcaacat	3300
acggaatate	atctacgatg	regragating	tteteece	actacatttt	ccaattetta	3360
tattttgcca	gtaattatga	atototagge	ttcatctttc	atctgatgac	togatacaga	3420
ttgggccgac	gatgtagagg	ecetyette	stogagggta	taacaddtad	attaaaaaa	3480
atccatccat	tggaggtcag	adattgcatt	attaggggca	acceategt	tgattgactc	3540
catgtaagct	tcgggactaa	cctggaagat	greaggergg	gtgaactctt	cctgaatctc	3600
attacaaagt	aaatcaggtg	aggagggrgg	atactocast	tttgtggacc	aatcaaaggg	3660
aggaaaaagc	ttatttgcag	agracicada	ttatttaaaa	ategaggaca	aaataatoto	3720
gagetette	tggatcatgg	agaggtacte	ttcctttacc	tetgaateag	attttcctag	3780
tegeattatt	tcatctttag	tananataca	teteteaac	acadecadad	attecttgag	3840
gaagggggac	ttcctaggaa	tgaaagcacc	gggagtgtga	atatttagat	gaaacccatt	3900
aatgtaatee	ctcactctgt aaccttgagt	caactgactt	tategeette	tetagetgaa	gcaatgcatg	3960
tatatcaaag	cttccatctt	tatatacata	tecageacat	agaatatatt	toggaatcca	4020
taaatgcaaa	agctcccaga	tatgtgttt	agagette	ggattttctg	gacactttgg	4080
acgaacgacg	aacgtgttag	catteetata	tgagaactga	coattagata	aggaggaggc	4140
ataggttagg	aacgigilag	greetaaaa	gagaacega	taggagetee	aaactctata	4200
catageegae	gegeettega	gaggergagg	tecattetet	tatagtggtt	gtaaatgggc	4260
gtatacccgt	gegeettega	aaccegeege	gegegeacta	atattaccoc	gccttcttt	4320
eggaceggge	cggcccagca	gyaaaayaay	ctttgattta	aagcctggtt	ctgctttgta	4380
cctgcgaggg	cooggggrag	astotasaca	aaccogtccc	aggcactata	aattgcctaa	4440
cgatttattt	tasttasta	atcotttees	ctcoagtcta	agaacccaa	ttcgccctat	4500
caagtgcgat	ccarccargg	actorcosts	dtttacaac	atcataacta	ggaaaaccct	4560
agrgagregr	actacaattc	actiggoogto	catececett	teaceaacte	gcgtaatagc	4620
ggegttacee	additableg	cocttoccas	cadttococo	geetataegt	acggcagttt	4680
gaagaggccc	gcaccyaccy	agagagaga	tatogtotot	ttgtggatgt	acagagtgat	4740
aaggtttaca	coccacaaay	agagageege	atccccta	ccagtgcace	tctgctgtca	4800
attattgaca	cgccggggcg	acggarggrg	accountings		, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	

WO 01/77350